

Heuristic algorithms for finding distribution reducts in probabilistic rough set model

Xi'ao Ma^{a,*}, Guoyin Wang^b, Hong Yu^b

^a*School of Computer & Information Engineering, Zhejiang Gongshang University, Hangzhou, 310018, P.R. China*

^b*Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing, 400065, P.R. China*

Abstract

Attribute reduction is one of the most important topics in rough set theory. Heuristic attribute reduction algorithms have been presented to solve the attribute reduction problem. It is generally known that fitness functions play a key role in developing heuristic attribute reduction algorithms. The monotonicity of fitness functions can guarantee the validity of heuristic attribute reduction algorithms. In probabilistic rough set model, distribution reducts can ensure the decision rules derived from the reducts are compatible with those derived from the original decision table. However, there are few studies on developing heuristic attribute reduction algorithms for finding distribution reducts. This is partly due to the fact that there are no monotonic fitness functions that are used to design heuristic attribute reduction algorithms in probabilistic rough set model. The main objective of this paper is to develop heuristic attribute reduction algorithms for finding distribution reducts in probabilistic rough set model. For one thing, two monotonic fitness functions are constructed, from which equivalence definitions of distribution reducts can be obtained. For another, two modified monotonic fitness functions are proposed to evaluate the significance of attributes more effectively. On this basis, two heuristic attribute reduction algorithms for finding distribution reducts are developed based on addition-deletion method and deletion method. In particular, the monotonicity of fitness functions guarantees the rationality of the proposed heuristic attribute reduction algorithms. Results of experimental analysis are included to quantify the effectiveness of the proposed fitness functions and distribution reducts.

Keywords: Attribute reduction; heuristic attribute reduction algorithm; the significance of attribute; probabilistic rough set model; distribution reduct.

*Corresponding author. Tel: +86 13588030950

Email addresses: maxiao73559@163.com (Xi'ao Ma), wanggy@ieee.org (Guoyin Wang), hongyu.cqpt@gmail.com (Hong Yu)

1. Introduction

Rough set theory, introduced by Pawlak [37], is a valid mathematical theory that deals well with imprecise, vague and uncertain information, and it has become an area of active research spreading throughout many fields, such as machine learning, data mining, knowledge discovery, intelligent data analyzing [5, 10, 11, 24, 40, 44]. In rough set theory, attribute reduction is a popular task to select the essential attributes that preserve or improve a certain classification property as the entire set of available attributes. Hence, an attribute reduct can be defined as a minimal attribute set that can preserve or improve specific classification criterion of a given information system. Attribute reduction is often helpful to reduce the computational cost, save storage space, improve learning performance and prevent over-fitting [48]. Studies on attribute reduction in rough set theory can be mainly divided into two categories. The first category concentrates on the study of definition of attribute reduct. The second category focuses on the study of attribute reduction algorithm.

For definition of attribute reduct, one mainly emphasizes on selecting what kinds of properties of a given information system to keep unchanged or to improve [15, 33, 34]. For example, Pawlak [38] defined a relative reduct that keeps the quality of classification or classification positive region unchanged. Miao et al. [31] constructed the mutual information based reducts to extract relevant attribute sets that preserve the mutual information of a given decision table. Kryszkiewicz [18, 19] investigated and compared the relationships among possible reduct, approximate reduct, μ -reduct, μ -decision reduct and generalized decision reduct. Zhang et al. [58] introduced the maximum distribution reduct, which preserves all maximum decision rules in a decision table. Deng et al. [7] proposed the notions of conditional knowledge granularity to reflect the relationship between conditional attributes and decision attribute, and defined an attribute reduct based on conditional knowledge granularity. Jiang et al. [17] proposed a new model of relative decision entropy by combining roughness with the degree of dependency, and used it as the reduction criterion.

For attribute reduction algorithm, one mainly focuses on designing the effective reduct construction methods for finding a specific type of reduct [45, 47, 62]. According to the different methods, research efforts on attribute reduction algorithm can be divided into three categories. The first class of methods is based on discernibility matrix. Miao et al. [32] discussed the structures of discernibility matrices for three different reducts, including region preservation reduct, decision preservation reduct and relationship preservation reduct. Yao and Zhao [55] introduced a reduct construction method based on discernibility matrix simplification. The second class of methods is based on heuristics. Qian et al. [42] proposed the concept of positive approximation for accelerating heuristic attribute reduction algorithms. Parthala et al. [35, 36] proposed a distance measure-based attribute reduction algorithm by considering the proximity of objects in the boundary region to those in the lower approximation. The third class of methods is based on stochastic optimization. Chen et al. [4] proposed a novel attribute reduction algorithm based on ant colony optimization for finding

a reduct that keeps the mutual information unchanged. Ye et al. [57] introduced a novel fitness function for designing the particle swarm optimization-based and genetic-based attribute reduction algorithms.

In particular, the studies mentioned above are mainly focused on attribute reduction in classical rough set model. Relative to the classical rough set model, probabilistic rough set model allows certain acceptable levels of errors by using the threshold parameters [8, 26, 28, 49]. Hence, the probabilistic rough set model can effectively deal with data sets which have the noisy and uncertain data. By setting the different threshold parameters, one can derive many existing probabilistic rough set models, such as 0.5 probabilistic rough set model [39], decision-theoretic rough set model [52], variable precision rough set model [63], Bayesian rough set model [46] and game rough set model [12]. Although attribute reduction in probabilistic rough set model has gained considerable attention in recently, these works generally focus on the study of definition of attribute reduct [13, 14, 21, 30, 50, 54, 59, 61, 63]. For attribute reduction algorithm, the research efforts mainly concentrate on the discernibility matrix based methods [13, 30, 51, 61] and stochastic optimization based methods [16, 27, 43]. Few attempts have been made to study the heuristic attribute reduction algorithms in probabilistic rough set model. This is partly because the attribute reduction in probabilistic rough set model becomes more complex after introducing the threshold parameters.

This paper concentrates on constructing heuristic attribute reduction algorithms for finding low and upper distribution reducts in probabilistic rough set model, which are defined by Mi et al. in [30]. Generally speaking, the most common heuristic attribute reduction algorithms are the addition-deletion method and the deletion method [56]. The addition-deletion method starts from an empty set or the core, then it adds attributes one by one on the basis of the significance of attributes until a stopping criterion is reached. On the contrary, the deletion method starts with a set containing all the attributes, then it delete the attributes one by one according to the significance of attributes. In fact, the heuristic attribute reduction algorithms mainly included two aspects: the stopping criterion and the significance measures of attributes. The stopping criterion is implemented by checking the jointly sufficient condition in the definition of attribute reduct, and the significance measures of attributes is used to rank the attributes. As we known, the fitness functions play a non-trivial role in designing stopping criteria and constructing significance measures of attributes. Furthermore, the monotonicity of fitness functions is very important to guarantee the validity of heuristic attribute reduction algorithms. However, very little work have considered the monotonicity of fitness functions in attribute reduction of probabilistic rough set model. In this paper, we first construct two monotonic fitness functions in probabilistic rough set model. The equivalent definitions of the low and upper distribution reducts are obtained based on the constructed monotonic fitness functions. Then we use them to design the stopping criterion of heuristic attribute reduction algorithms. The monotonicity of fitness functions guarantees the validity of heuristic attribute reduction algorithms. In addition, we propose two modified monotonic fitness functions

to evaluate the significance of attributes more effectively. On this basis, we develop two heuristic attribute reduction algorithms for finding the low and upper distribution reducts based on addition-deletion method and deletion method. In the end, some experimental analyses are covered to validate the effectiveness of the proposed fitness functions and distribution reducts.

The remaining sections of this paper are organized as follows. Section 2 briefly reviews some basic notions related to the definition of distribution reducts and heuristic attribute reduction algorithms. Section 3 develops two heuristic attribute reduction algorithms for finding the low and upper distribution reducts in probabilistic rough set model by constructing the monotonic fitness functions. Several experiments are given to illustrate the effectiveness of the constructed fitness functions and distribution reducts in Section 4. Section 5 concludes and offers suggestions for future research.

2. Preliminary knowledge

In this section, we recall the basic notions related to attribute reduction on rough set theory. [25, 52].

2.1. The definition of distribution reducts

An information system is a four-tuple $IS = (U, A, V, f)$, where U is a finite nonempty set of objects called universe, A is a nonempty finite set of attributes, $V = \bigcup_{a \in A} V_a$, where V_a is a nonempty set of values of attribute $a \in A$, called the domain of a , $f : U \times A \rightarrow V$ is a mapping that maps an object in U to exactly one value in V_a such that $\forall a \in A, x \in U, f(x, a) \in V_a$. For brevity, $IS = (U, A, V, f)$ can be written as $IS = (U, A)$.

For any subset of attributes $R \subseteq A$, an indiscernibility relation $IND(R)$ on U is defined as:

$$IND(R) = \{(x, y) \in U \times U \mid \forall a \in R, f(x, a) = f(y, a)\}.$$

It can be easily shown that $IND(R)$ is an equivalence relation on U . For $R \subseteq A$, the equivalence relation $IND(R)$ partitions U into some equivalence classes denoted by $U/IND(R) = \{[x]_R \mid x \in U\}$, for simplicity, $U/IND(R)$ will be replaced by U/R , where $[x]_R$ is an equivalence class determined by x with respect to R , i.e., $[x]_R = \{y \in U \mid (x, y) \in IND(R)\}$.

To describe a concept, rough set theory introduces a pair of lower approximation and upper approximation as follows.

Definition 1. Given an information system $IS = (U, A)$, $R \subseteq A$ and $X \subseteq U$, the lower approximation and upper approximation of X with respect to R are defined as:

$$\begin{aligned} \underline{apr}_R(X) &= \{x \in U \mid [x]_R \subseteq X\}, \\ \overline{apr}_R(X) &= \{x \in U \mid [x]_R \cap X \neq \emptyset\}. \end{aligned}$$

The lower approximation of X is a set of objects that belong to X with certainty, and the upper approximation of X is a set of objects that possibly belong to X .

Given an information system $IS = (U, A)$, $P, Q \subseteq A$, one can define a partial relation \preceq on 2^A as follows [6]:

$$P \preceq Q \Leftrightarrow \forall x \in U, [x]_P \subseteq [x]_Q.$$

If $P \preceq Q$, Q is said to be coarser than P (or P is finer than Q). If $P \preceq Q$ and $P \neq Q$, Q is said to be strictly coarser than P or P is strictly finer than Q , denoted by $P \prec Q$. In fact, $P \prec Q \Leftrightarrow \forall x \in U$, we have that $[x]_P \subseteq [x]_Q$ and there exists $y \in U$ such that $[y]_P \subset [y]_Q$.

A decision table is a four-tuple $DT = (U, C \cup D, V, f)$, where C is condition attribute set, D is decision attribute set, and $C \cap D = \emptyset$, V is the union of attribute domain, $V = V_C \cup V_D = \{V_a | a \in C\} \cup \{V_d | d \in D\}$. $DT = (U, C \cup D, V, f)$ can be written as $DT = (U, C \cup D)$ more simply.

Definition 2. [22] *Given a decision table $DT = (U, C \cup D)$, $R \subseteq C$ and $U/D = \{Y_1, Y_2, \dots, Y_M\}$ is a classification of the universe U . The positive region of U/D with respect to R is defined as follows:*

$$POS_R(D) = \bigcup_{1 \leq i \leq M} \underline{apr}_R(Y_i).$$

Pawlak rough set model does not allow any tolerance of errors. The probabilistic rough set model, which is a main extension of Pawlak rough set model, shows certain levels of tolerance for errors.

Definition 3. *Given an information system $IS = (U, A)$, for any $0 \leq \beta < \alpha \leq 1$, $R \subseteq A$ and $X \subseteq U$. The probabilistic lower approximation and probabilistic upper approximation of X with respect to R are defined as follows:*

$$\begin{aligned} \underline{apr}_R^{(\alpha, \beta)}(X) &= \{x \in U | p(X|[x]_R) \geq \alpha\}, \\ \overline{apr}_R^{(\alpha, \beta)}(X) &= \{x \in U | p(X|[x]_R) > \beta\}, \end{aligned}$$

where $p(X|[x]_R) = \frac{|[x]_R \cap X|}{|[x]_R|}$.

Definition 4. [22] *Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$, $R \subseteq C$ and $U/D = \{Y_1, Y_2, \dots, Y_M\}$ is a classification of the universe U . The probabilistic positive region of U/D with respect to R is defined as follows:*

$$POS_R^{(\alpha, \beta)}(D) = \bigcup_{1 \leq i \leq M} \underline{apr}_R^{(\alpha, \beta)}(Y_i).$$

Attribute reduction is one of the most important topics in rough set theory. An attribute reduct is defined as a subset of attributes that are jointly sufficient and individually necessary for preserving or improving a particular property of the given information system [54]. A general definition of an attribute reduct is given as follows.

Definition 5. [60] Given an information system $IS = (U, A)$, $R \subseteq A$ and consider a certain property \mathbb{P} , which can be represented by an evaluation function $e : 2^A \rightarrow (L, \preceq)$, of IS . An attribute set R is called a reduct of IS if it satisfies the following two conditions:

- (1) Jointly sufficient condition: $e(A) \preceq e(R)$,
- (2) Individually necessary condition: for any $R' \subset R$, $\neg(e(A) \preceq e(R'))$.

An evaluation or fitness function, $e : 2^A \rightarrow (L, \preceq)$, maps an attribute set to an element of a poset L equipped with the partial order relation \preceq i.e., \preceq is reflexive, anti-symmetric and transitive. For a certain property \mathbb{P} , various fitness functions can be used to evaluate the degree of satisfiability of the property by an attribute set. Generally, the fitness function is not unique. The jointly sufficient condition guarantees that the evaluation of the reduct R with respect to e is the same or superior to $e(A)$, and it has $e(R) = e(A)$ in many cases. The individually necessary condition guarantees that the reduct is minimal, namely, there is no redundant or superfluous attribute in the reduct R .

According to the different properties \mathbb{P} of an information system, the different reducts can be defined. There are a huge amount of known properties, such as a description of an object relation [37], partitions of an information system [60], a classification of a set of concepts [38], where the classification of a set of concepts is the most common property in rough set theory.

In classical rough set model, the classification of a set of concepts can be evaluated by using the positive region of the classification (Definition 2). In probabilistic rough set model, the classification of a set of concepts can be evaluated by using the probabilistic positive region of the classification (Definition 4). However, the decision rules derived from the reduct preserving probabilistic positive region maybe in conflict with those derived from the original decision table because of non-monotonicity of probabilistic positive region with respect to the set inclusion of attribute sets [30].

To derive the conflict free decision rules, Mi et al. [30] presented the concepts of distribution reducts based on variable precision rough set model which is a typical probabilistic rough set model.

Definition 6. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$, $R \subseteq C$ and $U/D = \{Y_1, Y_2, \dots, Y_M\}$ is a classification of the universe U . The (α, β) lower and upper approximation distribution functions with respect to R are defined as:

$$\begin{aligned} \underline{apr}_R^{(\alpha, \beta)} &= (\underline{apr}_R^{(\alpha, \beta)}(Y_1), \underline{apr}_R^{(\alpha, \beta)}(Y_2), \dots, \underline{apr}_R^{(\alpha, \beta)}(Y_M)), \\ \overline{apr}_R^{(\alpha, \beta)} &= (\overline{apr}_R^{(\alpha, \beta)}(Y_1), \overline{apr}_R^{(\alpha, \beta)}(Y_2), \dots, \overline{apr}_R^{(\alpha, \beta)}(Y_M)). \end{aligned}$$

The (α, β) lower or upper approximation distribution functions can be seen as fitness functions for evaluating the classification of a set of concepts in probabilistic rough set model. By the (α, β) lower and upper approximation distribution functions, the (α, β) lower and upper approximation distribution reducts based on probabilistic rough set model are defined as follows.

Definition 7. [30] Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$ and $R \subseteq C$, we have

- (1) If $\underline{apr}_R^{(\alpha, \beta)} = \underline{apr}_C^{(\alpha, \beta)}$, R is referred to as an (α, β) lower distribution consistent set of DT ; if $\underline{apr}_R^{(\alpha, \beta)} = \underline{apr}_C^{(\alpha, \beta)}$ and $\underline{apr}_{R'}^{(\alpha, \beta)} \neq \underline{apr}_C^{(\alpha, \beta)}$ for all $R' \subset R$, then R is referred to as an (α, β) lower distribution reduct of DT .
- (2) If $\overline{apr}_R^{(\alpha, \beta)} = \overline{apr}_C^{(\alpha, \beta)}$, R is referred to as an (α, β) upper distribution consistent set of DT ; if $\overline{apr}_R^{(\alpha, \beta)} = \overline{apr}_C^{(\alpha, \beta)}$ and $\overline{apr}_{R'}^{(\alpha, \beta)} \neq \overline{apr}_C^{(\alpha, \beta)}$ for all $R' \subset R$, then R is referred to as an (α, β) upper distribution reduct of DT .

An (α, β) lower (upper) distribution reduct is a minimal subset of attribute set that preserves the (α, β) lower (upper) approximations of all decision classes. For the sake of the simplicity, the (α, β) lower and upper distribution reducts are collectively called distribution reducts in the rest of this paper.

It is important to note that the monotonicity property of the fitness function used in the definition of attribute reduct with respect to the set inclusion of attribute sets should receive enough attention when we design the heuristic attribute reduction algorithms. If the fitness function is monotonic regarding the set inclusion of attribute sets, individually necessary condition only need to consider the subsets $R - \{a\}$ for all $a \in R$ to guarantee a reduct R is minimal. If the fitness function is not monotonic regarding the set inclusion of attribute sets, individually necessary condition must consider all subsets of a reduct R to make sure it is a minimal set [14, 61].

In Definition 7, individually necessary conditions must consider all subsets of the reduct R because the (α, β) lower and upper approximation distribution functions are not monotonic regarding the set inclusion of attribute sets, which will complicate the algorithm design.

2.2. Typical heuristic attribute reduction algorithms

So far, Yao et al. [56] have summarized three groups of heuristic attribute reduction algorithms based on the addition-deletion method, the deletion method and the addition method, where the addition-deletion method based algorithm and the deletion method based algorithm are two most widely used heuristic attribute reduction algorithms by the rough set community. Hence, we mainly discuss the first two methods based heuristic attribute reduction algorithms in this subsection.

The addition-deletion method based algorithm and the deletion method based algorithm are displayed in Algorithm 1 and Algorithm 2, respectively.

Algorithm 1 starts with an empty set or the core, and consequently adds attributes to the subset of selected attributes until a candidate reduct satisfies the jointly sufficient condition in the definition of attribute reduct. Each selected attribute maximizes the increment of fitness values of the current attribute subset. One needs the deleting process to delete the superfluous attributes in the candidate reduct one by one after the addition process because the addition process may add the superfluous attributes.

Algorithm 1 The addition-deletion method for computing a reduct

Input: A decision table $DT = (U, C \cup D)$, threshold (α, β)

Output: A reduct R

Method: Addition-deletion method

```

1: // Addition
2: Let  $R = \emptyset$ ,  $CA = C$ 
3: while  $R$  is not jointly sufficient and  $CA \neq \emptyset$  do
4:   Compute fitness values of all the attributes in  $CA$  regarding the property
      $\mathbb{P}$  using a fitness function  $\sigma$ 
5:   Select an attribute  $c$  according to its fitness, let  $CA = CA - \{c\}$ 
6:   Let  $R = R \cup \{c\}$ 
7: end while
8: // Deletion
9: Let  $CD = R$ 
10: while  $CD \neq \emptyset$  do
11:   Compute fitness values of all the attributes in  $CD$  regarding the property
      $\mathbb{P}$  using a fitness function  $\delta$ 
12:   Select an attribute  $a$  according to its fitness, let  $CD = CD - \{a\}$ 
13:   if  $R - \{a\}$  is jointly sufficient then
14:      $R = R - \{a\}$ 
15:   end if
16: end while
17: Return  $R$ 

```

Algorithm 2 The deletion method for computing a reduct

Input: A decision table $DT = (U, C \cup D)$, threshold (α, β)

Output: A reduct R

Method: Deletion method

```

1: // Deletion
2: Let  $R = C$ ,  $CD = C$ 
3: while  $CD \neq \emptyset$  do
4:   Compute fitness values of all the attributes in  $CD$  regarding the property
      $\mathbb{P}$  using a fitness function  $\delta$ 
5:   Select an attribute  $a$  according to its fitness, let  $CD = CD - \{a\}$ 
6:   if  $R - \{a\}$  is jointly sufficient then
7:      $R = R - \{a\}$ 
8:   end if
9: end while
10: Return  $R$ 

```

Algorithm 2 takes the entire condition attributes as a candidate reduct, then selects the attributes for deleting one by one according to the fitness values. If

the subset of the remaining attributes satisfies the jointly sufficient condition in the definition of attribute reduct after deleting the selected attribute, then the attribute is the superfluous attribute and can be deleted. A reduct is obtained if and only if each attribute has been checked once.

Algorithm 1 and Algorithm 2 mainly consist of two key steps: checking the jointly sufficient condition and evaluating the significance of attributes. Checking the jointly sufficient condition is to ensure that the candidate reduct R meet jointly sufficient condition in the definition of attribute reduct. Evaluating the significance of attributes is to sort attributes and provide the heuristic information for searching a reduct, and the step can be implemented by designing the effective fitness functions. Different fitness functions may get the different orders of attributes, that may obtain the different reducts.

It is worth pointing out that the monotonicity of the fitness function in Definition 5 with respect to the set inclusion of attribute sets is very important for the completeness of Algorithm 1 and Algorithm 2 [29]. If the fitness function is monotonic regarding the set inclusion of attribute sets, Algorithm 1 and Algorithm 2 must obtain a reduct, namely, Algorithm 1 and Algorithm 2 are complete. If the fitness function is not monotonic regarding the set inclusion of attribute sets, Algorithm 1 and Algorithm 2 may obtain a super reduct that includes redundancy attributes, namely, Algorithm 1 and Algorithm 2 are incomplete. Moreover, the fitness functions for evaluating the significance of attributes should also satisfy the monotonicity with respect to the set inclusion of attribute sets and provide enough precision to sort the attributes more effectively [17, 20].

3. Heuristic algorithms to find distribution reducts in probabilistic rough set model

Checking the jointly sufficient condition and evaluating the significance of attributes are two key steps in heuristic attribute reduction algorithms based on the addition-deletion strategy and the deletion strategy. Furthermore, the monotonicity of the fitness functions for checking the jointly sufficient condition and evaluating the significance of attributes is very important for the validity of the heuristic attribute reduction algorithms. To obtain the distribution reducts with heuristic attribute reduction algorithms, in this section, we first construct two monotonic fitness functions. Then we give the equivalent definition of distribution reducts based on the monotonic fitness functions constructed. After that, we further proposed two significance measures of attributes by multiplying measures of granularity of partitions by the monotonic fitness functions constructed. Moreover, the core and core computation algorithm for distribution reducts are also presented. On this basis, two heuristic attribute reduction algorithms to find distribution reducts are developed based on addition-deletion method and deletion method. Finally, an illustrative example to the heuristic attribute reduction algorithms proposed is provided step by step.

3.1. The equivalent definition of distribution reducts with monotonic fitness functions

For a certain property in the definition of attribution reduct, the different fitness functions can be used as its indicator. The monotonicity of the fitness functions is very important to guarantee the completeness of heuristic attribute reduction algorithms. To develop the complete heuristic attribute reduction algorithms to obtain distribution reducts, we construct two monotonic fitness functions in this subsection. The equivalent definition of distribution reducts are further given based on the monotonic fitness functions constructed.

Now, let us first give two monotonic fitness functions as follow.

Definition 8. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$, $R \subseteq C$ and $U/D = \{Y_1, Y_2, \dots, Y_M\}$ is a classification of the universe U , we denote

$$\eta_R^{(\alpha, \beta)} = \frac{\sum_{Y_i \in U/D} |\underline{apr}_R(\underline{apr}_C^{(\alpha, \beta)}(Y_i))|}{|U||U/D|},$$

$$\mu_R^{(\alpha, \beta)} = \frac{\sum_{Y_i \in U/D} |\overline{apr}_R(\overline{apr}_C^{(\alpha, \beta)}(Y_i))|}{|U||U/D|}.$$

If $R = \emptyset$, then define $\eta_R^{(\alpha, \beta)} = 0$ and $\mu_R^{(\alpha, \beta)} = 1$. Moreover, $\eta_R^{(\alpha, \beta)}$ and $\mu_R^{(\alpha, \beta)}$ are denoted by η and μ respectively if there is no confusion arisen.

It is easy to see from above that $\eta_R^{(\alpha, \beta)}$ and $\mu_R^{(\alpha, \beta)}$ are defined by using the lower approximations of all elements in $\underline{apr}_C^{(\alpha, \beta)}$ with respect R and upper approximations of all elements in $\overline{apr}_C^{(\alpha, \beta)}$ with respect to R , respectively. $\underline{apr}_R(\underline{apr}_C^{(\alpha, \beta)}(Y_i))$ represents the change of certain knowledge with respect to $\underline{apr}_C^{(\alpha, \beta)}(Y_i)$ after removing attributes $C - R$ from the decision table, and $\overline{apr}_R(\overline{apr}_C^{(\alpha, \beta)}(Y_i))$ represents the change of relevant knowledge with respect to $\overline{apr}_C^{(\alpha, \beta)}(Y_i)$ after removing attributes $C - R$ from the decision table. Therefore, $\eta_R^{(\alpha, \beta)}$ represents the change of certain knowledge with respect to $\underline{apr}_C^{(\alpha, \beta)}$. $\mu_R^{(\alpha, \beta)}$ represents the change of relevant knowledge with respect to $\overline{apr}_C^{(\alpha, \beta)}$. Hence, we can use $\eta_R^{(\alpha, \beta)}$ and $\mu_R^{(\alpha, \beta)}$ to compute the (α, β) lower and upper distribution reducts, respectively.

Theorem 1. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$ and $P, Q \subseteq C$, we have

$$(1) P \preceq Q \Rightarrow \eta_P^{(\alpha, \beta)} \geq \eta_Q^{(\alpha, \beta)},$$

$$(2) P \preceq Q \Rightarrow \mu_P^{(\alpha, \beta)} \leq \mu_Q^{(\alpha, \beta)}.$$

Proof. (1) Suppose $P \preceq Q$. In terms of the definition of \preceq , we have $[x]_P \subseteq [x]_Q$. Let $U/D = \{Y_1, Y_2, \dots, Y_M\}$ and $Y_i \in U/D, 1 \leq i \leq M$.

On the one hand, for $\forall x \in \underline{apr}_Q(\underline{apr}_C^{(\alpha, \beta)}(Y_i))$, one can obtain $[x]_Q \subseteq \underline{apr}_C^{(\alpha, \beta)}(Y_i)$, hence $[x]_P \subseteq \underline{apr}_C^{(\alpha, \beta)}(Y_i)$, then we obtain that $x \in \underline{apr}_P(\underline{apr}_C^{(\alpha, \beta)}(Y_i))$, thus $\underline{apr}_P(\underline{apr}_C^{(\alpha, \beta)}(Y_i)) \supseteq \underline{apr}_Q(\underline{apr}_C^{(\alpha, \beta)}(Y_i))$.

In the other hand, for $\forall x \in \overline{apr}_P(\overline{apr}_C^{(\alpha, \beta)}(Y_i))$, we have $[x]_P \cap \overline{apr}_C^{(\alpha, \beta)}(Y_i) \neq \emptyset$. Since $[x]_P \subseteq [x]_Q$, $[x]_Q \cap \overline{apr}_C^{(\alpha, \beta)}(Y_i) \neq \emptyset$, we have $x \in \overline{apr}_Q(\overline{apr}_C^{(\alpha, \beta)}(Y_i))$. It follows that $\overline{apr}_P(\overline{apr}_C^{(\alpha, \beta)}(Y_i)) \subseteq \overline{apr}_Q(\overline{apr}_C^{(\alpha, \beta)}(Y_i))$.

As a result, we have

$$\begin{aligned} |\underline{apr}_P(\underline{apr}_C^{(\alpha, \beta)}(Y_i))| &\geq |\underline{apr}_Q(\underline{apr}_C^{(\alpha, \beta)}(Y_i))|, \\ |\overline{apr}_P(\overline{apr}_C^{(\alpha, \beta)}(Y_i))| &\leq |\overline{apr}_Q(\overline{apr}_C^{(\alpha, \beta)}(Y_i))|. \end{aligned}$$

Thus,

$$\begin{aligned} \sum_{Y_i \in U/D} |\underline{apr}_P(\underline{apr}_C^{(\alpha, \beta)}(Y_i))| &\geq \sum_{Y_i \in U/D} |\underline{apr}_Q(\underline{apr}_C^{(\alpha, \beta)}(Y_i))|, \\ \sum_{Y_i \in U/D} |\overline{apr}_P(\overline{apr}_C^{(\alpha, \beta)}(Y_i))| &\leq \sum_{Y_i \in U/D} |\overline{apr}_Q(\overline{apr}_C^{(\alpha, \beta)}(Y_i))|. \end{aligned}$$

Hence,

$$\begin{aligned} \eta_P^{(\alpha, \beta)} &= \frac{\sum_{Y_i \in U/D} |\underline{apr}_P(\underline{apr}_C^{(\alpha, \beta)}(Y_i))|}{|U||U/D|} \geq \eta_Q^{(\alpha, \beta)} = \frac{\sum_{Y_i \in U/D} |\underline{apr}_Q(\underline{apr}_C^{(\alpha, \beta)}(Y_i))|}{|U||U/D|}, \\ \mu_P^{(\alpha, \beta)} &= \frac{\sum_{Y_i \in U/D} |\overline{apr}_P(\overline{apr}_C^{(\alpha, \beta)}(Y_i))|}{|U||U/D|} \leq \mu_Q^{(\alpha, \beta)} = \frac{\sum_{Y_i \in U/D} |\overline{apr}_Q(\overline{apr}_C^{(\alpha, \beta)}(Y_i))|}{|U||U/D|}. \end{aligned}$$

This completes the proof.

The proof of (2) is similar to that of (1).

By Theorem 1 we immediately get the following corollary.

Corollary 1. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$ and $P, Q \subseteq C$, we have

- (1) $P \supseteq Q \Rightarrow \eta_P^{(\alpha, \beta)} \geq \eta_Q^{(\alpha, \beta)}$,
- (2) $P \supseteq Q \Rightarrow \mu_P^{(\alpha, \beta)} \leq \mu_Q^{(\alpha, \beta)}$.

Theorem 1 and Corollary 1 show that the fitness function $\eta_R^{(\alpha, \beta)}$ increases and the fitness function $\mu_R^{(\alpha, \beta)}$ decreases as the equivalence classes become smaller through finer partitioning, which means that adding a new attribute into the existing subset of condition attributes at least does not decrease $\eta_R^{(\alpha, \beta)}$ or increase $\mu_R^{(\alpha, \beta)}$, and that deleting an attribute from the existing subset of condition attributes at least does not increase $\eta_R^{(\alpha, \beta)}$ or decreases $\mu_R^{(\alpha, \beta)}$. The property is very important for constructing heuristic attribute reduction algorithms.

In the following, the performance of Theorem 1 is shown through an illustrative example.

Example 1. Given a decision table $DT = (U, C \cup D)$ showed in Table 1, where $U = \{x_1, x_2, \dots, x_{12}\}$, and $C = \{a_1, a_2, a_3, a_4, a_5, a_6\}$. Suppose that $\alpha = 0.60$ and $\beta = 0.40$, $P, Q \subseteq C$, where $P = \{a_1, a_2, a_3\}$ and $Q = \{a_1, a_2\}$. As we can see, $P \supseteq Q$, which means $P \preceq Q$.

Table 1: A decision table

U	a_1	a_2	a_3	a_4	a_5	a_6	d
x_1	1	1	1	1	0	0	0
x_2	0	1	1	0	1	0	0
x_3	0	1	1	1	0	0	0
x_4	0	1	1	1	0	0	1
x_5	0	0	0	1	0	1	0
x_6	0	0	1	1	1	1	1
x_7	0	0	1	1	0	1	1
x_8	1	1	0	0	1	1	0
x_9	1	1	0	0	1	1	1
x_{10}	1	1	0	0	0	0	1
x_{11}	1	1	0	0	0	0	1

By calculating, one can have

$$\begin{aligned} U/C &= \{\{x_1\}, \{x_2\}, \{x_3, x_4\}, \{x_5\}, \{x_6\}, \{x_7\}, \{x_8, x_9\}, \{x_{10}, x_{11}\}\}, \\ U/D &= \{\{x_1, x_2, x_3, x_5, x_8\}, \{x_4, x_6, x_7, x_9, x_{10}, x_{11}\}\}. \end{aligned}$$

Hence, we have

$$\begin{aligned} \underline{apr}_C^{(0.60, 0.40)} &= (\{x_1, x_2, x_5\}, \{x_6, x_7, x_{10}, x_{11}\}), \\ \overline{apr}_C^{(0.60, 0.40)} &= (\{x_1, x_2, x_3, x_4, x_5, x_8, x_9\}, \{x_3, x_4, x_6, x_7, x_8, x_9, x_{10}, x_{11}\}). \end{aligned}$$

It can be easily calculated that

$$\begin{aligned} U/P &= \{\{x_1\}, \{x_2, x_3, x_4\}, \{x_5\}, \{x_6, x_7\}, \{x_8, x_9, x_{10}, x_{11}\}\}, \\ U/Q &= \{\{x_1\}, \{x_2, x_3, x_4\}, \{x_5, x_6, x_7\}, \{x_8, x_9, x_{10}, x_{11}\}\}. \end{aligned}$$

Obviously, $P \prec Q$.

According to Definition 8, we have

$$\begin{aligned} \eta_P^{(0.60, 0.40)} &= \frac{\sum_{Y_i \in U/D} |\underline{apr}_P(\underline{apr}_C^{(0.60, 0.40)}(Y_i))|}{|U||U/D|} = \frac{2+2}{11 \times 2} = \frac{4}{22}, \\ \eta_Q^{(0.60, 0.40)} &= \frac{\sum_{Y_i \in U/D} |\underline{apr}_Q(\underline{apr}_C^{(0.60, 0.40)}(Y_i))|}{|U||U/D|} = \frac{1+0}{11 \times 2} = \frac{1}{22}, \\ \mu_P^{(0.60, 0.40)} &= \frac{\sum_{Y_i \in U/D} |\overline{apr}_P(\overline{apr}_C^{(0.60, 0.40)}(Y_i))|}{|U||U/D|} = \frac{9+9}{11 \times 2} = \frac{18}{22}, \\ \mu_Q^{(0.60, 0.40)} &= \frac{\sum_{Y_i \in U/D} |\overline{apr}_Q(\overline{apr}_C^{(0.60, 0.40)}(Y_i))|}{|U||U/D|} = \frac{11+10}{11 \times 2} = \frac{21}{22}. \end{aligned}$$

Thus, $\eta_P^{(0.60, 0.40)} > \eta_Q^{(0.60, 0.40)}$ and $\mu_P^{(0.60, 0.40)} < \mu_Q^{(0.60, 0.40)}$. It is clear that $\eta_R^{(\alpha, \beta)}$ increases and $\mu_R^{(\alpha, \beta)}$ decreases with R becoming finer.

Now, let us give equivalence descriptions for each distribution consistent set by the fitness functions $\eta_R^{(\alpha, \beta)}$ and $\mu_R^{(\alpha, \beta)}$.

Theorem 2. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$ and $R \subseteq C$, we have

- (1) R is an (α, β) lower distribution consistent set of DT iff $\eta_R^{(\alpha, \beta)} = \eta_C^{(\alpha, \beta)}$,
- (2) R is an (α, β) upper distribution consistent set of DT iff $\mu_R^{(\alpha, \beta)} = \mu_C^{(\alpha, \beta)}$.

Proof. Since $R \subseteq C$, it is easy to verify that $\xi([x]_R) = \{[y]_C : [y]_C \subseteq [x]_R\}$ forms a partition of $[x]_R$. Let $U/D = \{Y_1, Y_2, \dots, Y_M\}$.

(1) “ \Rightarrow ” Since R is an (α, β) lower distribution consistent set, we have $\underline{apr}_R^{(\alpha, \beta)} = \underline{apr}_C^{(\alpha, \beta)}$, therefore for $\forall Y_i \in U/D$, we obtain that $\underline{apr}_R^{(\alpha, \beta)}(Y_i) = \underline{apr}_C^{(\alpha, \beta)}(Y_i)$.
Hence,

$$\begin{aligned} \eta_R^{(\alpha, \beta)} &= \frac{\sum_{Y_i \in U/D} |\underline{apr}_R^{(\alpha, \beta)}(Y_i)|}{|U||U/D|} = \frac{\sum_{Y_i \in U/D} |\underline{apr}_C^{(\alpha, \beta)}(Y_i)|}{|U||U/D|} \\ &= \frac{\sum_{Y_i \in U/D} |\underline{apr}_R^{(\alpha, \beta)}(Y_i)|}{|U||U/D|} = \frac{\sum_{Y_i \in U/D} |\underline{apr}_C^{(\alpha, \beta)}(Y_i)|}{|U||U/D|} \\ &= \frac{\sum_{Y_i \in U/D} |\underline{apr}_C^{(\alpha, \beta)}(Y_i)|}{|U||U/D|} = \eta_C^{(\alpha, \beta)}. \end{aligned}$$

“ \Leftarrow ” For $\forall Y_i \in U/D$, we have $\underline{apr}_R^{(\alpha, \beta)}(Y_i) \subseteq \underline{apr}_C^{(\alpha, \beta)}(Y_i)$.

Suppose there exists $Y_i \in U/D$ such that $\underline{apr}_R^{(\alpha, \beta)}(Y_i) \subset \underline{apr}_C^{(\alpha, \beta)}(Y_i)$.
As a result,

$$\begin{aligned} \eta_R^{(\alpha, \beta)} &= \frac{\sum_{Y_i \in U/D} |\underline{apr}_R^{(\alpha, \beta)}(Y_i)|}{|U||U/D|} < \frac{\sum_{Y_i \in U/D} |\underline{apr}_C^{(\alpha, \beta)}(Y_i)|}{|U||U/D|} \\ &= \frac{\sum_{Y_i \in U/D} |\underline{apr}_C^{(\alpha, \beta)}(Y_i)|}{|U||U/D|} = \eta_C^{(\alpha, \beta)}. \end{aligned}$$

It conflicts with condition $\eta_R^{(\alpha, \beta)} = \eta_C^{(\alpha, \beta)}$. Hence, for $\forall Y_i \in U/D$, we have $\underline{apr}_R^{(\alpha, \beta)}(Y_i) = \underline{apr}_C^{(\alpha, \beta)}(Y_i)$.

If $x \in \underline{apr}_C^{(\alpha, \beta)}(Y_i)$, there are two cases in which $[x]_R \subseteq \underline{apr}_C^{(\alpha, \beta)}(Y_i)$ and $[x]_R \not\subseteq \underline{apr}_C^{(\alpha, \beta)}(Y_i)$.

When $[x]_R \not\subseteq \underline{apr}_C^{(\alpha, \beta)}(Y_i)$, we have $x \notin \underline{apr}_R^{(\alpha, \beta)}(Y_i)$.

It conflicts with condition $\underline{apr}_R^{(\alpha, \beta)}(Y_i) = \underline{apr}_C^{(\alpha, \beta)}(Y_i)$. Thus $[x]_R \subseteq \underline{apr}_C^{(\alpha, \beta)}(Y_i)$. Since $[x]_R = \cup\{[y]_C : [y]_C \in \xi([x]_R)\}$, we obtain that $[y]_C \subseteq$

$[x]_R \subseteq \underline{apr}_C^{(\alpha, \beta)}(Y_i)$ for all $[y]_C \in \xi([x]_R)$. That is to say, for all $[y]_C \in \xi([x]_R)$, it has $p(Y_i|[y]_C) \geq \alpha$. Therefore we have that

$$\begin{aligned} p(Y_i|[x]_R) &= \left(\sum \{ |[y]_C \cap Y_i| : [y]_C \in \xi([x]_R) \} \right) / |[x]_R| \\ &= \sum \left\{ p(Y_i|[y]_C) \cdot \frac{|[y]_C|}{|[x]_R|} : [y]_C \in \xi([x]_R) \right\} \\ &\geq \alpha \sum \left\{ \frac{|[y]_C|}{|[x]_R|} : [y]_C \in \xi([x]_R) \right\} = \alpha. \end{aligned}$$

As a result $x \in \underline{apr}_R^{(\alpha, \beta)}(Y_i)$.

On the other hand, if $x \in \underline{apr}_R^{(\alpha, \beta)}(Y_i)$, then we have $[x]_R \subseteq \underline{apr}_R^{(\alpha, \beta)}(Y_i)$. Since $\underline{apr}_R(\underline{apr}_C^{(\alpha, \beta)}(Y_i)) = \underline{apr}_C^{(\alpha, \beta)}(Y_i)$, there are two cases in which $[x]_R \subseteq \underline{apr}_C^{(\alpha, \beta)}(Y_i)$ and $[x]_R \cap \underline{apr}_C^{(\alpha, \beta)}(Y_i) = \emptyset$.

When $[x]_R \cap \underline{apr}_C^{(\alpha, \beta)}(Y_i) = \emptyset$, since $[x]_R = \cup \{ [y]_C : [y]_C \in \xi([x]_R) \}$, we have $[y]_C \cap \underline{apr}_C^{(\alpha, \beta)}(Y_i) = \emptyset$ for all $[y]_C \in \xi([x]_R)$. That is to say, for all $[y]_C \in \xi([x]_R)$, it has $p(Y_i|[y]_C) < \alpha$. Therefore we have that

$$\begin{aligned} p(Y_i|[x]_R) &= \left(\sum \{ |[y]_C \cap Y_i| : [y]_C \in \xi([x]_R) \} \right) / |[x]_R| \\ &= \sum \left\{ p(Y_i|[y]_C) \cdot \frac{|[y]_C|}{|[x]_R|} : [y]_C \in \xi([x]_R) \right\} \\ &< \alpha \sum \left\{ \frac{|[y]_C|}{|[x]_R|} : [y]_C \in \xi([x]_R) \right\} = \alpha. \end{aligned}$$

As a result $[x]_R \cap \underline{apr}_R^{(\alpha, \beta)}(Y_i) = \emptyset$, which contradicts with $[x]_R \subseteq \underline{apr}_R^{(\alpha, \beta)}(Y_i)$.

Hence $[x]_R \subseteq \underline{apr}_C^{(\alpha, \beta)}(Y_i)$, and in tune $x \in \underline{apr}_C^{(\alpha, \beta)}(Y_i)$.

Thus we conclude that $\underline{apr}_R^{(\alpha, \beta)}(Y_i) = \underline{apr}_C^{(\alpha, \beta)}(Y_i)$ for $\forall Y_i \in U/D$, i.e., R is an (α, β) lower distribution consistent set.

The proof of (2) is similar to that of (1).

Theorem 3. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$ and $R \subseteq C$, we have

- (1) An attribute $a \in R$ is dispensable in R with respect to $\underline{apr}_R^{(\alpha, \beta)}$ iff $\eta_{R-\{a\}}^{(\alpha, \beta)} = \eta_R^{(\alpha, \beta)}$,
- (2) An attribute $a \in R$ is dispensable in R with respect to $\overline{apr}_R^{(\alpha, \beta)}$ iff $\mu_{R-\{a\}}^{(\alpha, \beta)} = \mu_R^{(\alpha, \beta)}$.

By Theorem 3 we immediately get the following corollary.

Corollary 2. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$ and $R \subseteq C$, we have

- (1) An attribute $a \in R$ is indispensable in R with respect to $\underline{apr}_R^{(\alpha, \beta)}$ iff $\eta_{R-\{a\}}^{(\alpha, \beta)} \neq \eta_R^{(\alpha, \beta)}$,
- (2) An attribute $a \in R$ is indispensable in R with respect to $\overline{apr}_R^{(\alpha, \beta)}$ iff $\mu_{R-\{a\}}^{(\alpha, \beta)} \neq \mu_R^{(\alpha, \beta)}$.

Theorem 2 and Corollary 2 yield the following theorem.

Theorem 4. *Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$ and $R \subseteq C$, we have*

- (1) R is an (α, β) lower distribution reduct of DT iff
 - (I) $\eta_R^{(\alpha, \beta)} = \eta_C^{(\alpha, \beta)}$;
 - (II) $\eta_{R-\{a\}}^{(\alpha, \beta)} \neq \eta_C^{(\alpha, \beta)}$ for $\forall a \in R$.
- (2) R is an (α, β) upper distribution reduct of DT iff
 - (I) $\mu_R^{(\alpha, \beta)} = \mu_C^{(\alpha, \beta)}$;
 - (II) $\mu_{R-\{a\}}^{(\alpha, \beta)} \neq \mu_C^{(\alpha, \beta)}$ for $\forall a \in R$.

Theorem 4 gives the equivalent definition of distribution reducts. Compared with Definition 7, individually necessary conditions only need to check the subsets $R - \{a\}$ for all $a \in R$, not all subsets of R because the fitness functions $\eta_R^{(\alpha, \beta)}$ and $\mu_R^{(\alpha, \beta)}$ are monotonic with respect to the set inclusion of attribute sets. The condition simplifies the algorithm design.

In fact, Theorem 4 provides concrete methods to design heuristic algorithms for obtaining the distribution reducts, and guarantees the completeness of heuristic attribute reduction algorithms.

3.2. The significance measures of attributes

Evaluating the significance of attributes is one of the most important problem to design the efficient heuristic attribute reduction algorithms. In this subsection, we construct the significance measures of attributes on the basis of the monotonic fitness functions proposed in order to provide the heuristic information that can guide search to distribution reducts.

For the significance measures of attributes, the monotonicity of measures are very important. The monotonicity can guarantee the rationality of the measures to evaluate the significance of attributes. Hence, we can use the fitness functions $\eta_R^{(\alpha, \beta)}$ and $\mu_R^{(\alpha, \beta)}$ as the significance measures of attributes for searching the (α, β) lower and upper distribution reducts to a certain degree, respectively.

The corresponding significance measures of attributes are defined as follows.

Definition 9. *Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$, $R \subseteq C$ and $\forall a \in C - R$, the significance measures of attribute a in R with respect to $\underline{apr}_C^{(\alpha, \beta)}$ and $\overline{apr}_C^{(\alpha, \beta)}$ are defined as follows:*

- (1) $SIG_\eta(a, R, \underline{apr}_C^{(\alpha, \beta)}) = \eta_{R \cup \{a\}}^{(\alpha, \beta)} - \eta_R^{(\alpha, \beta)}$.

$$(2) \text{ } SIG_\mu(a, R, \overline{apr}_C^{(\alpha, \beta)}) = \mu_R^{(\alpha, \beta)} - \mu_{R \cup \{a\}}^{(\alpha, \beta)}.$$

Definition 9 can be used to provide heuristics to guide the mechanism of searching an attribute. $SIG_\eta(a, R, \underline{apr}_C^{(\alpha, \beta)})$ can serve as the heuristic information for searching the (α, β) lower distribution reduct. $SIG_\mu(a, R, \overline{apr}_C^{(\alpha, \beta)})$ can serve as the heuristic information for searching the (α, β) upper distribution reduct. For convenience, the corresponding significance measures of attributes are also denoted as $SIG_\eta(a, \underline{apr}_C^{(\alpha, \beta)}) = \eta_{\{a\}}^{(\alpha, \beta)}$ and $SIG_\mu(a, \overline{apr}_C^{(\alpha, \beta)}) = 1 - \mu_{\{a\}}^{(\alpha, \beta)}$ for any singleton attribute $a \in C$.

As to the monotonicity of the fitness functions $\eta_R^{(\alpha, \beta)}$ and $\mu_R^{(\alpha, \beta)}$, they can be used for evaluating the significance of attributes. However, the accuracy of fitness functions is very important to effectively evaluate the significance of attributes. In some situations, the fitness functions $\eta_R^{(\alpha, \beta)}$ and $\mu_R^{(\alpha, \beta)}$ cannot supply enough information for evaluating, because they do not taking into full account the granularity of partitions. The limitations are revealed by the following example.

Example 2. Given a decision table $DT = (U, C \cup D)$ showed in Table 2, where $U = \{x_1, x_2, \dots, x_{12}\}$, and $C = \{a_1, a_2, a_3, a_4, a_5, a_6\}$. Suppose that $\alpha = 0.60$ and $\beta = 0.40$, $P, Q \subseteq C$, where $P = \{a_1, a_2, a_3, a_4\}$ and $Q = \{a_1, a_2, a_3\}$. As we can see, $P \supseteq Q$, which means $P \preceq Q$.

Table 2: A decision table

U	a_1	a_2	a_3	a_4	a_5	a_6	d
x_1	1	0	0	1	1	0	0
x_2	0	0	1	0	1	1	0
x_3	0	1	1	0	1	0	0
x_4	0	1	1	0	1	0	1
x_5	0	1	1	0	1	0	1
x_6	0	1	1	1	0	0	1
x_7	1	0	0	1	0	1	0
x_8	1	0	0	1	0	1	1
x_9	1	0	0	0	1	1	0
x_{10}	1	0	0	0	1	0	0
x_{11}	1	0	0	0	1	0	1
x_{12}	1	0	0	0	1	0	1

By calculating, one can have

$$\begin{aligned} U/C &= \{\{x_1\}, \{x_2\}, \{x_3, x_4, x_5\}, \{x_6\}, \{x_7, x_8\}, \{x_9\}, \{x_{10}, x_{11}, x_{12}\}\}, \\ U/D &= \{\{x_1, x_2, x_3, x_7, x_9, x_{10}\}, \{x_4, x_5, x_6, x_8, x_{11}, x_{12}\}\}. \end{aligned}$$

Hence, we have

$$\begin{aligned}\underline{apr}_C^{(0.60,0.40)} &= (\{x_1, x_2, x_{10}, x_{11}, x_{12}\}, \{x_3, x_4, x_5, x_6, x_9\}), \\ \overline{apr}_C^{(0.60,0.40)} &= (\{x_1, x_2, x_7, x_8, x_{10}, x_{11}, x_{12}\}, \{x_3, x_4, x_5, x_6, x_7, x_8, x_9\}).\end{aligned}$$

It can be easily calculated that

$$\begin{aligned}U/P &= \{\{x_1, x_7, x_8\}, \{x_2\}, \{x_3, x_4, x_5\}, \{x_6\}, \{x_9, x_{10}, x_{11}, x_{12}\}\}, \\ U/Q &= \{\{x_1, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}\}, \{x_2\}, \{x_3, x_4, x_5, x_6\}\}.\end{aligned}$$

Obviously, $P \prec Q$.

According to Definition 8, we have

$$\begin{aligned}\eta_P^{(0.60,0.40)} &= \frac{\sum_{Y_i \in U/D} |\underline{apr}_P(\underline{apr}_C^{(0.60,0.40)}(Y_i))|}{|U||U/D|} = \frac{1+4}{12 \times 2} = \frac{5}{24}, \\ \eta_Q^{(0.60,0.40)} &= \frac{\sum_{Y_i \in U/D} |\underline{apr}_Q(\underline{apr}_C^{(0.60,0.40)}(Y_i))|}{|U||U/D|} = \frac{1+4}{12 \times 2} = \frac{5}{24}, \\ \mu_P^{(0.60,0.40)} &= \frac{\sum_{Y_i \in U/D} |\overline{apr}_P(\overline{apr}_C^{(0.60,0.40)}(Y_i))|}{|U||U/D|} = \frac{8+11}{12 \times 2} = \frac{19}{24}, \\ \mu_Q^{(0.60,0.40)} &= \frac{\sum_{Y_i \in U/D} |\overline{apr}_Q(\overline{apr}_C^{(0.60,0.40)}(Y_i))|}{|U||U/D|} = \frac{8+11}{12 \times 2} = \frac{19}{24}.\end{aligned}$$

As a result, we have

$$\eta_P^{(0.60,0.40)} = \eta_Q^{(0.60,0.40)} \text{ and } \mu_P^{(0.60,0.40)} = \mu_Q^{(0.60,0.40)}.$$

From Example 2, we can see that there is a partial relation between P and Q . However, we obtained the same values of the fitness functions. It indicates that the fitness functions in Definition 8 can not discern the attribute subsets P and Q clearly. The main reason is because they do not take into full account the granularity of partitions. Hence, it is easy to obtain that the significance of some attributes is zero by using the fitness functions in Definition 8. Therefore, it is necessary to introduce more effective fitness functions to evaluate the significance of attributes.

In what follows, we introduce more effective fitness functions based on the measures of granularity of partitions. We review several existing measures of granularity of partitions before introducing the more efficient fitness functions.

Given a partition, there are many methods that can be used to measure the granularity of a partition. Currently, Yao and Zhao [53] provided a unified framework for measures of granularity of partitions by considering the expected granularity of blocks in a partition.

First, let us recall the measure of granularity of a set.

Definition 10. [53] Suppose U is finite and nonempty universe. A function $m : 2^U \rightarrow \mathbb{R}$ is called a measure of granularity of a set if it satisfies the following conditions: for all $X, Y \in 2^U$.

- (1) (nonnegativity) $m(X) \geq 0$,
- (2) (monotonicity) $X \subset Y \Rightarrow m(X) < m(Y)$,
- (3) (size invariant) $X \equiv_s Y \Leftrightarrow m(X) = m(Y)$,

where the binary relation \equiv_s is defined by: $A \equiv_s B$ if there exists a bijection from A to B .

The measure of granularity of a partition is defined as follows based on the measure of granularity of a set.

Definition 11. [53] Suppose $\pi = \{X_1, X_2, \dots, X_K\}$ is a partition of a finite nonempty universe U and $m : 2^U \rightarrow \mathbb{R}$ is a measure of granularity of subsets of U , satisfies (1) – (3) of Definition 10. The expected granularity of blocks of π is defined as:

$$EG_m(\pi) = \mathbf{E}_{P_\pi}(m(\cdot)) = \sum_{i=1}^K m(X_i)p(X_i),$$

where $P_\pi = (p(X_1), p(X_2), \dots, p(X_K)) = \left(\frac{|X_1|}{|U|}, \frac{|X_2|}{|U|}, \dots, \frac{|X_K|}{|U|}\right)$ is the probability distribution defined by π and $\mathbf{E}_{P_\pi}(\cdot)$ is the mathematical expectation with respect to distribution P_π .

Let Π be the set of all partitions of U . The expected granularity $EG_m(\pi)$ reaches the minimum value if and only if $\pi = \Pi_0 = \{\{x\} | x \in U\}$ which is the finest partition in Π , and it reaches the maximum value if and only if $\pi = \Pi_1 = \{U\}$ which is the coarsest partition in Π . In general, we have $EG_m(\Pi_0) \leq EG_m(\pi) \leq EG_m(\Pi_1)$ for any $\pi \in \Pi$.

The expected granularity $EG_m(\pi)$ is a class of measures of granularity of partitions. The different measures of granularity of partitions can be derived from the expected granularity by considering various classes of measures of granularity of sets. Hence, many existing measures of granularity of partitions are instances of the expected granularity. The three most widely used measures of granularity of partitions are co-entropy, knowledge granularity and combination granularity in all existing measures of granularity of partitions. The corresponding definitions are as follows.

Definition 12. [2, 3, 22] Given an information system $IS = (U, A)$, $R \subseteq A$ and $U/R = \{X_1, X_2, \dots, X_K\}$. The co-entropy of U/R is defined as follows:

$$CE(U/R) = \sum_{i=1}^K \frac{|X_i|}{|U|} \log_2 |X_i|.$$

Obviously, one has that $0 \leq CE(U/R) \leq \log_2 |U|$.

Definition 13. [22, 23] Given an information system $IS = (U, A)$, $R \subseteq A$ and $U/R = \{X_1, X_2, \dots, X_K\}$. The knowledge granulation of U/R is defined as follows:

$$GK(U/R) = \frac{1}{|U|^2} \sum_{i=1}^K |X_i|^2.$$

Obviously, one has that $1/|U| \leq GK(U/R) \leq 1$.

Definition 14. [41] Given an information system $IS = (U, A)$, $R \subseteq A$ and $U/R = \{X_1, X_2, \dots, X_K\}$. The combination granulation of U/R is defined as follows:

$$CG(U/R) = \sum_{i=1}^K \frac{|X_i|}{|U|} \frac{\binom{|X_i|}{2}}{\binom{|U|}{2}}.$$

Obviously, one has that $0 \leq CG(U/R) \leq 1$.

In the following, we modified the fitness functions $\eta_R^{(\alpha, \beta)}$ and $\mu_R^{(\alpha, \beta)}$ in Definition 8 based on the expected granularity to evaluate the significance of attributes more effectively. Moreover, an example is provided to show that the modified fitness functions are more effective and suitable for evaluating the significance of attributes.

Definition 15. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$, $R \subseteq C$ and $U/D = \{Y_1, Y_2, \dots, Y_M\}$ is a classification of the universe U , we denote

$$\begin{aligned} G\eta_R^{(\alpha, \beta)} &= EG_m(\Pi_1) - (1 - \eta_R^{(\alpha, \beta)})EG_m(U/R), \\ G\mu_R^{(\alpha, \beta)} &= \mu_R^{(\alpha, \beta)}EG_m(U/R). \end{aligned}$$

If $R = \emptyset$, then define $G\eta_R^{(\alpha, \beta)} = 0$ and $G\mu_R^{(\alpha, \beta)} = EG_m(\Pi_1)$. Moreover, $G\eta_R^{(\alpha, \beta)}$ and $G\mu_R^{(\alpha, \beta)}$ are denoted by $G\eta$ and $G\mu$ respectively if there is no confusion arisen.

Clearly, $G\eta_R^{(\alpha, \beta)}$ and $G\mu_R^{(\alpha, \beta)}$ take into account the granularity of the partition.

Theorem 5. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$ and $P, Q \subseteq C$, we have

$$(1) P \preceq Q \Rightarrow G\eta_P^{(\alpha, \beta)} \geq G\eta_Q^{(\alpha, \beta)},$$

$$(2) P \preceq Q \Rightarrow G\mu_P^{(\alpha, \beta)} \leq G\mu_Q^{(\alpha, \beta)}.$$

Proof. It can be easily proved according to Theorem 1, Definition 10 and Definition 11.

By Theorem 5 we immediately get the following corollary.

Corollary 3. *Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$ and $P, Q \subseteq C$, we have*

- (1) $P \supseteq Q \Rightarrow G\eta_P^{(\alpha, \beta)} \geq G\eta_Q^{(\alpha, \beta)}$,
- (2) $P \supseteq Q \Rightarrow G\mu_P^{(\alpha, \beta)} \leq G\mu_Q^{(\alpha, \beta)}$.

Theorem 5 and Corollary 3 show that the modified fitness function $G\eta_R^{(\alpha, \beta)}$ increases and the modified fitness function $G\mu_R^{(\alpha, \beta)}$ decreases as the equivalence classes become smaller through finer partitioning, which means that adding a new attribute into the existing subset of condition attributes at least does not decrease $G\eta_R^{(\alpha, \beta)}$ or increase $G\mu_R^{(\alpha, \beta)}$, and that deleting an attribute from the existing subset of condition attributes at least does not increase $G\eta_R^{(\alpha, \beta)}$ or decrease $G\mu_R^{(\alpha, \beta)}$.

Example 3. *Continued from Example 2.*

We take the co-entropy (Definition 12) as an example of the expected granularity (Definition 11).

According to Definition 12, we have

$$\begin{aligned} EG_m(U/P) &= \frac{3}{12}\log_2 3 + \frac{1}{12}\log_2 1 + \frac{3}{12}\log_2 3 + \frac{1}{12}\log_2 1 + \frac{4}{12}\log_2 4 \\ &\approx 1.46, \\ EG_m(U/Q) &= \frac{7}{12}\log_2 7 + \frac{1}{12}\log_2 1 + \frac{4}{12}\log_2 4 \\ &\approx 2.30. \end{aligned}$$

According to Definition 15, we have

$$\begin{aligned} G\eta_P^{(0.60, 0.40)} &= EG_m(\Pi_1) - \eta_P^{(0.60, 0.40)} EG_m(U/P) = \log_2 12 - \frac{5}{23} \times 1.46 \approx 3.27, \\ G\eta_Q^{(0.60, 0.40)} &= EG_m(\Pi_1) - \eta_Q^{(0.60, 0.40)} EG_m(U/Q) = \log_2 12 - \frac{5}{23} \times 2.30 \approx 3.08, \\ G\mu_P^{(0.60, 0.40)} &= EG_m(\Pi_1) - \mu_P^{(0.60, 0.40)} EG_m(U/P) = \frac{19}{24} \times 1.46 \approx 1.16, \\ G\mu_Q^{(0.60, 0.40)} &= EG_m(\Pi_1) - \mu_Q^{(0.60, 0.40)} EG_m(U/Q) = \frac{19}{24} \times 2.30 \approx 1.82. \end{aligned}$$

Obviously, $G\eta_P^{(0.60, 0.40)} > G\eta_Q^{(0.60, 0.40)}$ and $G\mu_P^{(0.60, 0.40)} < G\mu_Q^{(0.60, 0.40)}$. It is clear that the modified fitness function $G\eta_R^{(\alpha, \beta)}$ increases and the modified fitness function $G\mu_R^{(\alpha, \beta)}$ decreases with R becoming finer.

Example 3 shows the modified fitness functions $G\eta_R^{(\alpha, \beta)}$ and $G\mu_R^{(\alpha, \beta)}$ are more powerful for evaluating the attribute subsets in some cases. Hence, we use them as the significance measures of attributes for guiding search to the (α, β) lower and upper distribution reducts, respectively.

The corresponding significance measures of attributes are defined as follows.

Definition 16. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$, $R \subset C$ and $\forall a \in C - R$, the significance measures of attribute a in R with respect to $\underline{apr}_C^{(\alpha, \beta)}$ and $\overline{apr}_C^{(\alpha, \beta)}$ are defined as follows:

- (1) $SIG_{G\eta}(a, R, \underline{apr}_C^{(\alpha, \beta)}) = G\eta_{R \cup \{a\}}^{(\alpha, \beta)} - G\eta_R^{(\alpha, \beta)}.$
- (2) $SIG_{G\mu}(a, R, \overline{apr}_C^{(\alpha, \beta)}) = G\mu_R^{(\alpha, \beta)} - G\mu_{R \cup \{a\}}^{(\alpha, \beta)}.$

Similar to Definition 9, for convenience, the corresponding significance measures of attributes are also denoted as $SIG_{G\eta}(a, \underline{apr}_C^{(\alpha, \beta)}) = \eta_{\{a\}}^{(\alpha, \beta)}$ and $SIG_{G\mu}(a, \overline{apr}_C^{(\alpha, \beta)}) = EG_m(\Pi_1) - \mu_{\{a\}}^{(\alpha, \beta)}$ for any singleton attribute $a \in C$.

3.3. The attribute core

Attribute core plays important role in heuristic attribute reduction algorithms. The attribute core is the set of all indispensable attributes. In other words, the attribute core is included in all reducts. Hence, it is often selected as the starting point in heuristic attribute reduction algorithms to narrow the search space of attributes. Moreover, the different attribute cores can be defined according to the different definitions of attribute reducts. In this subsection, we define the attribute cores for distribution reducts and provide the method for calculating the attribute cores.

Definition 17. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$, the attribute cores for the (α, β) lower and upper distribution reducts are defined as follows:

- (1) $CORE_{\underline{apr}^{(\alpha, \beta)}}(C) = \cap RED_{\underline{apr}^{(\alpha, \beta)}}(C),$
- (2) $CORE_{\overline{apr}^{(\alpha, \beta)}}(C) = \cap RED_{\overline{apr}^{(\alpha, \beta)}}(C).$

where, $RED_{\underline{apr}^{(\alpha, \beta)}}(C)$ and $RED_{\overline{apr}^{(\alpha, \beta)}}(C)$ denote the set of all (α, β) lower and (α, β) upper distribution reducts, respectively.

Theorem 2 and Definition 17 yield the following theorem.

Theorem 6. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$ and $c \in C$, we have

- (1) $c \in CORE_{\underline{apr}^{(\alpha, \beta)}}(C)$ iff $\eta_{C - \{c\}}^{(\alpha, \beta)} \neq \eta_C^{(\alpha, \beta)},$
- (2) $c \in CORE_{\overline{apr}^{(\alpha, \beta)}}(C)$ iff $\mu_{C - \{c\}}^{(\alpha, \beta)} \neq \mu_C^{(\alpha, \beta)}.$

Proof. (1) “ \Rightarrow ” Suppose $\eta_{C - \{c\}}^{(\alpha, \beta)} = \eta_C^{(\alpha, \beta)}$. c is dispensable in C according to Theorem 3. It conflicts with condition $c \in CORE_{\underline{apr}^{(\alpha, \beta)}}(C)$, so we have $\eta_{C - \{c\}}^{(\alpha, \beta)} \neq \eta_C^{(\alpha, \beta)}.$

- “ \Leftarrow ” If $\eta_{C-\{c\}}^{(\alpha,\beta)} \neq \eta_C^{(\alpha,\beta)}$, then c is indispensable in C according to Corollary 2. Hence, c must appear in all (α, β) lower distribution reducts of DT . Thus, $c \in \cap RED_{\underline{apr}^{(\alpha,\beta)}}(C)$, i.e., $c \in CORE_{\underline{apr}^{(\alpha,\beta)}}(C)$.
- (2) The proof is similar to that of (1).

Theorem 6 and Definition 17 yield the following definition.

Definition 18. Given a decision table $DT = (U, C \cup D)$, for any $0 \leq \beta < \alpha \leq 1$, the attribute cores for the (α, β) lower and upper distribution reducts are defined as follows:

- (1) $CORE_{\underline{apr}^{(\alpha,\beta)}}(C) = \{c \in C \mid \eta_{C-\{c\}}^{(\alpha,\beta)} \neq \eta_C^{(\alpha,\beta)}\}$,
- (2) $CORE_{\overline{apr}^{(\alpha,\beta)}}(C) = \{c \in C \mid \mu_{C-\{c\}}^{(\alpha,\beta)} \neq \mu_C^{(\alpha,\beta)}\}$.

Definition 18 states the calculation method of the attribute cores for the (α, β) lower and upper distribution reducts.

An algorithm for calculating the attribute cores for the (α, β) lower and upper distribution reducts is displayed in Algorithm 3.

Algorithm 3 The algorithm for calculating the attribute core for the (α, β) lower (upper) distribution reduct

Input: A decision table $DT = (U, C \cup D)$, threshold values (α, β)

Output: An attribute core for the (α, β) lower (upper) distribution reduct

Note: $(T, \Delta) \in \{(\underline{apr}, \eta), (\overline{apr}, \mu)\}$

- 1: Let $CORE_{T^{(\alpha,\beta)}}(C) = \emptyset$
 - 2: **for** Each $c \in C$ **do**
 - 3: Calculate $\Delta_{C-\{c\}}^{(\alpha,\beta)}$
 - 4: **if** $\Delta_{C-\{c\}}^{(\alpha,\beta)} \neq \Delta_C^{(\alpha,\beta)}$ **then**
 - 5: $CORE_{T^{(\alpha,\beta)}}(C) = CORE_{T^{(\alpha,\beta)}}(C) \cup \{c\}$
 - 6: **end if**
 - 7: **end for**
 - 8: Return $CORE_{T^{(\alpha,\beta)}}(C)$
-

3.4. Attribute reduction algorithms

In this subsection, we proposed two heuristic attribute reduction algorithms based on the addition-deletion method (Algorithm 1) and the deletion method (Algorithm 2) to obtain the distribution reducts in probabilistic rough set model. Two algorithm details are shown in Algorithm 4 and Algorithm 5.

Algorithm 4 The addition-deletion method for computing the distribution reducts of DT

Input: A decision table $DT = (U, C \cup D)$, threshold values (α, β)

Output: An (α, β) lower (upper) distribution reduct of DT

Method: Addition-deletion method

Note: $(T, \Delta) \in \{(\underline{apr}, \eta), (\overline{apr}, \mu)\}$

```

1: Calculate  $CORE_{T(\alpha, \beta)}(C)$  by Algorithm 3
2: Calculate  $\Delta_C^{(\alpha, \beta)}$ 
3: Let  $R = CORE_{T(\alpha, \beta)}(C)$ ,  $CA = C - R$ 
4: if  $\Delta_R^{(\alpha, \beta)} = \Delta_C^{(\alpha, \beta)}$  then
5:   go to Step 28
6: end if
7: // Addition
8: while  $\Delta_R^{(\alpha, \beta)} \neq \Delta_C^{(\alpha, \beta)}$  do
9:   for Each  $a \in CA$  do
10:    Calculate  $SIG_{G\Delta}(a, R, T_C^{(\alpha, \beta)})$ 
11:   end for
12:   if  $SIG_{G\Delta}(a, R, T_C^{(\alpha, \beta)}) = \max_{a \in C-R} SIG_{G\Delta}(a, R, T_C^{(\alpha, \beta)})$  then
13:      $R = R \cup \{a\}$ ,  $CA = CA - \{a\}$ 
14:   end if
15: end while
16: // Deletion
17: Let  $CD = R$ 
18: for Each  $a \in CD$  do
19:   Calculate  $SIG_{G\Delta}(a, T)$ 
20: end for
21: Sort attributes in  $CD$  according to  $SIG_{G\Delta}(a, T)$  in a ascending order
22: while  $CD \neq \emptyset$  do
23:    $CD = CD - \{a\}$ , where  $a$  is the first element of  $CD$ 
24:   if  $\Delta_{R-\{a\}}^{(\alpha, \beta)} = \Delta_C^{(\alpha, \beta)}$  then
25:      $R = R - \{a\}$ 
26:   end if
27: end while
28: Return  $R$ 

```

Algorithm 5 The deletion method for computing the distribution reducts of DT

Input: A decision table $DT = (U, C \cup D)$, threshold values (α, β)

Output: An (α, β) lower (upper) distribution reduct of DT

Method: Deletion method

Note: $(T, \Delta) \in \{(\underline{apr}, \eta), (\overline{apr}, \mu)\}$

```

1: Calculate  $\Delta_C^{(\alpha, \beta)}$ 
2: Let  $R = C$ ,  $CD = C$ 
3: for Each  $a \in CD$  do
4:   Calculate  $SIG_{G\Delta}(a, T)$ 
5: end for
6: Sort attributes in  $CD$  according to  $SIG_{G\Delta}(a, T)$  in a ascending order
7: while  $CD \neq \emptyset$  do
8:    $CD = CD - \{a\}$ , where  $a$  is the first element of  $CD$ 
9:   if  $\Delta_{R-\{a\}}^{(\alpha, \beta)} = \Delta_C^{(\alpha, \beta)}$  then
10:     $R = R - \{a\}$ 
11:   end if
12: end while
13: Return  $R$ 

```

3.5. An illustrative example

We have developed two heuristic attribute reduction algorithms to obtain distribution reducts. In this subsection, we present an example to show the validity of the proposed algorithms.

Example 4. (Continued from Example 1) For Table 1 shown in Example 1, we take the $(0.6, 0.4)$ low distribution reduct as an example. Moreover, the co-entropy (Definition 12) is used as an example of the expected granularity (Definition 11). We can calculate the $(0.60, 0.40)$ low distribution reduct by Algorithm 4.

By computing, we can obtain $\eta_C^{(0.60, 0.40)} \approx 0.32$.

According to core computing step in Algorithm 4, we first calculate $\eta_{C-\{a_i\}}^{(0.60, 0.40)}$ for each attribute $a_i \in C$ as follows:

$$\eta_{C-\{a_1\}}^{(0.60, 0.40)} \approx 0.27, \eta_{C-\{a_2\}}^{(0.60, 0.40)} \approx 0.32, \eta_{C-\{a_3\}}^{(0.75, 0.60)} \approx 0.23,$$

$$\eta_{C-\{a_4\}}^{(0.60, 0.40)} \approx 0.32, \eta_{C-\{a_5\}}^{(0.60, 0.40)} \approx 0.32, \eta_{C-\{a_6\}}^{(0.75, 0.60)} \approx 0.32.$$

Therefore, $CORE_{\underline{apr}^{(0.60, 0.40)}}(C) = \{a_1, a_3\}$.

Let $R = CORE_{\underline{apr}^{(0.60, 0.40)}}(C) = \{a_1, a_3\}$, then $\eta_R^{(0.60, 0.40)} \approx 0.09$.

Because $\eta_R^{(0.60, 0.40)} \neq \eta_C^{(0.60, 0.40)}$, according to the addition step in Algorithm

4 we calculate

$$G\eta_{R \cup \{a_2\}}^{(0.60, 0.40)} \approx 1.64, G\eta_{R \cup \{a_4\}}^{(0.60, 0.40)} \approx 1.53,$$

$$G\eta_{R \cup \{a_5\}}^{(0.60, 0.40)} \approx 1.84, G\eta_{R \cup \{a_6\}}^{(0.60, 0.40)} \approx 1.91.$$

Because $G\eta_{R \cup \{a_6\}}^{(0.60, 0.40)}$ is maximum, we select a_6 . Let $R = R \cup \{a_6\} = \{a_1, a_3, a_6\}$, then $\eta_R^{(0.60, 0.40)} \approx 0.27$.

Because $\eta_R^{(0.60, 0.40)} \neq \eta_C^{(0.60, 0.40)}$, we calculate $\eta_{R \cup \{a_i\}}^{(0.60, 0.40)}$ for each attribute $a_i \in C - R$ as follows:

$G\eta_{R \cup \{a_2\}}^{(0.60, 0.40)} \approx 1.91$, $G\eta_{R \cup \{a_4\}}^{(0.60, 0.40)} \approx 2.05$, $G\eta_{R \cup \{a_5\}}^{(0.60, 0.40)} \approx 2.14$.

Because $G\eta_{R \cup \{a_5\}}^{(0.60, 0.40)}$ is maximum, we select a_5 . Let $R = R \cup \{a_5\} = \{a_1, a_3, a_5, a_6\}$, then $\eta_R^{(0.60, 0.40)} \approx 0.32$.

Thus, we have $\eta_R^{(0.60, 0.40)} = \eta_C^{(0.60, 0.40)}$ when $R = \{a_1, a_3, a_5, a_6\}$.

According to the deletion step in Algorithm 4, we get a $(0.60, 0.40)$ low distribution reduct $\{a_1, a_3, a_5, a_6\}$ because we have $\eta_{R - \{a_i\}}^{(0.60, 0.40)} \neq \eta_C^{(0.60, 0.40)}$ for $\forall a_i \in R$.

It can be easily calculated that

$$\underline{apr}_{\{a_1, a_3, a_5, a_6\}}^{(0.60, 0.40)} = (\{x_1, x_2, x_5\}, \{x_6, x_7, x_{10}, x_{11}\}).$$

Hence, we have $\underline{apr}_{\{a_1, a_3, a_5, a_6\}}^{(0.60, 0.40)} = \underline{apr}_C^{(0.60, 0.40)}$.

In other words, the attribute set $\{a_1, a_3, a_5, a_6\}$ can keep the $(0.60, 0.40)$ lower approximations of all decision classes unchanged.

Similar to the computation of the $(0.60, 0.40)$ low distribution reduct, it is not difficult to find that $\{a_1, a_3, a_4, a_6\}$ is a $(0.60, 0.40)$ upper distribution reduct.

By Algorithm 5 we can obtain $\{a_1, a_3, a_4, a_6\}$ is a $(0.60, 0.40)$ low distribution reduct and $\{a_1, a_2, a_3, a_5\}$ is a $(0.60, 0.40)$ upper distribution reduct.

4. Experimental results

In this section, a series of experiments were designed to demonstrate that our methods proposed are effective and applicable. Ten benchmark real-world data sets were chosen for experimental evaluation. All the data sets were obtained from the UCI Repository of Machine Learning databases [1]. These data sets have been widely used in literatures. The general information about the selected UCI data sets is summarized in Table 3, where $|U|$ and $|C|$ denote the number of objects and the condition attributes, respectively. $|V_d|$ denotes the number of decision classes.

Since the data sets may contain missing values or continuous attributes, they would be handled in advance prior to attribute reduction. Missing values were filled with mean values for continuous attributes and mode values for nominal attributes. Continuous attributes were discretized using equal-frequency discretization method. All preprocessing methods were implemented by using WEKA filters [9].

4.1. The monotonicity experiments

In this subsection, several experiments were performed to verify the effectiveness of the proposed fitness functions in Section 3. In the experiments, we took the co-entropy (CE, Definition 12), knowledge granulation (KG, Definition 13) and combination granulation (CG, Definition 14) as examples of the expected

Table 3: Description of the datasets

ID	Data sets	$ U $	$ C $	$ V_d $
1	Horse-colic	368	22	2
2	Primary-tumor	339	17	21
3	Vehicle	846	18	4
4	Voting	436	16	2
5	Zoo	101	16	7
6	Credit Approval	690	15	2
7	Segment	2310	19	7
8	Kr-vs-kp	3196	36	2
9	Hypothyroid	3772	29	4
10	German	1000	20	2

granularity. Hence, both of fitness functions $G\eta$ and $G\mu$ (Definition 15) can adopt three types of implementations based on CE, KG and CG. The threshold parameters α and β are set to 0.6 and 0.4 respectively.

Figures 1 - 10 present the experimental results of the proposed fitness functions on ten data sets. In each of figures, the X-axis represents the size of condition attribute subset. The condition attribute subset is increased from one attribute to all attributes during the experiments. The Y-axis pertains to values of fitness functions. Furthermore, each figure has two subfigures. The subfigure (a) shows that the experimental results of the fitness functions η , $G\eta - CE$, $G\eta - KG$ and $G\eta - CG$, where $G\eta - CE$, $G\eta - KG$ and $G\eta - CG$ represent the CE-based, KG-based and CG-based implementations of $G\eta$ respectively. The subfigure (b) shows that the experimental results of the fitness functions μ , $G\mu - CE$, $G\mu - KG$ and $G\mu - CG$, where $G\mu - CE$, $G\mu - KG$ and $G\mu - CG$ represent the CE-based, KG-based and CG-based implementations of $G\mu$ respectively. Moreover, we rescaled the values of them to the $[0, 1]$ range in order to better visualize the data because the values of the fitness functions $G\eta - CE$ and $G\mu - CE$ may be greater than 1.

It can be seen from these figures that the values of the fitness functions η , $G\eta - CE$, $G\eta - KG$ and $G\eta - CG$ increase with the number of selected attributes becoming bigger, and the values of the fitness functions μ , $G\mu - CE$, $G\mu - KG$ and $G\mu - CG$ decrease with the number of selected attributes becoming bigger. It indicates that the proposed fitness functions are monotonic with respect to the set inclusion of attributes. The results are consistent with Corollary 1 and Corollary 3. However, it is easy to see that the values of the fitness function η are the same when the number of attributes increased from 1 to 2 on all the data sets except Zoo. Similarly, there is no change in the results of the fitness function μ as the number of attributes increases from 1 to 2 on some data sets, such as Horse-colic, Voting and Kr-vs-kp. In these cases, the fitness functions η and μ can not evaluate the significance of attributes effectively. For the same situation, the values of the fitness functions $G\eta - CE$, $G\eta - KG$ and $G\eta - CG$ get

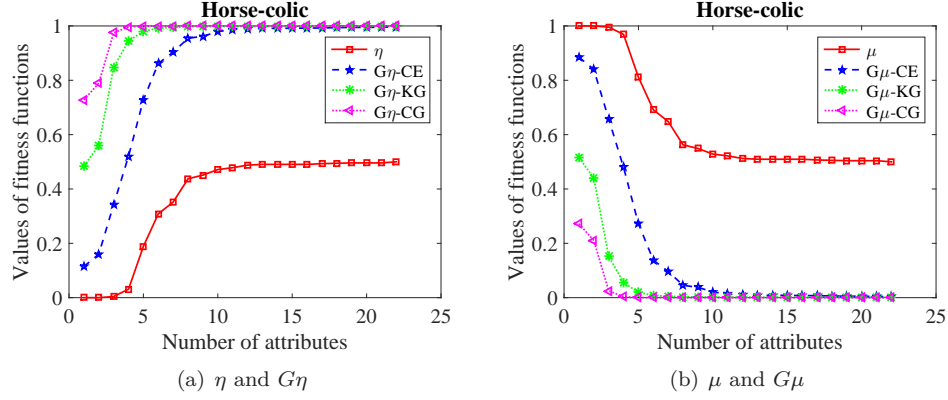


Figure 1: Monotonicity of fitness functions on data set Horse-colic.

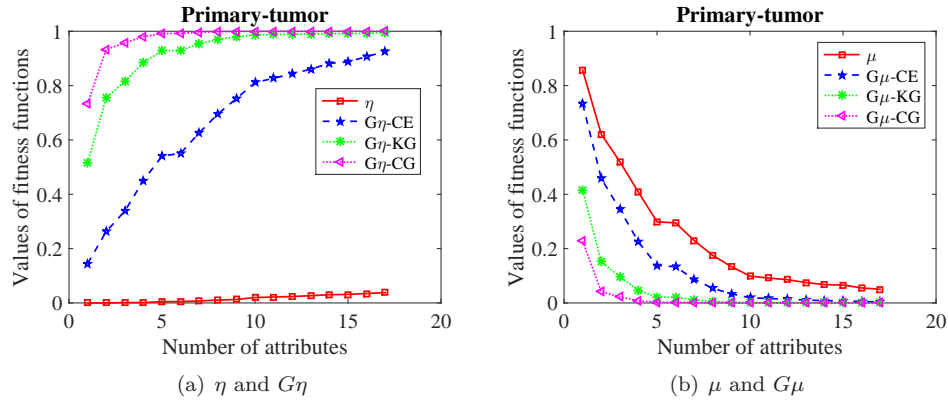


Figure 2: Monotonicity of fitness functions on data set Primary-tumor.

bigger, while the values of the fitness functions $G\mu-CE$, $G\mu-KG$ and $G\mu-CG$ get smaller when the number of attributes increased from 1 to 2. It shows that the fitness functions $G\eta$ and $G\mu$ can evaluate the significance of attributes more accurately. The results show that the fitness functions $G\eta$ and $G\mu$ can provide more information for evaluating the significance of attributes. In other words, the fitness functions $G\eta$ and $G\mu$ have a better discrimination power than the fitness functions η and μ . Hence, we can evaluate the significance of attributes more effectively by using the fitness functions $G\eta$ and $G\mu$.

4.2. Significance of single attributes

Evaluating single attributes and ranking them are an important step in Algorithms 4 and 5. Hence, in this subsection, we compared the effectiveness of the proposed fitness functions in evaluating the significance of single attributes.

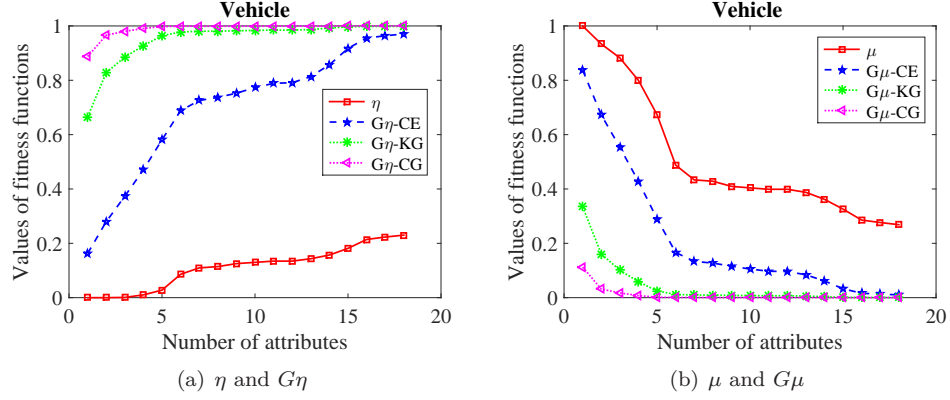


Figure 3: Monotonicity of fitness functions on data set Vehicle.

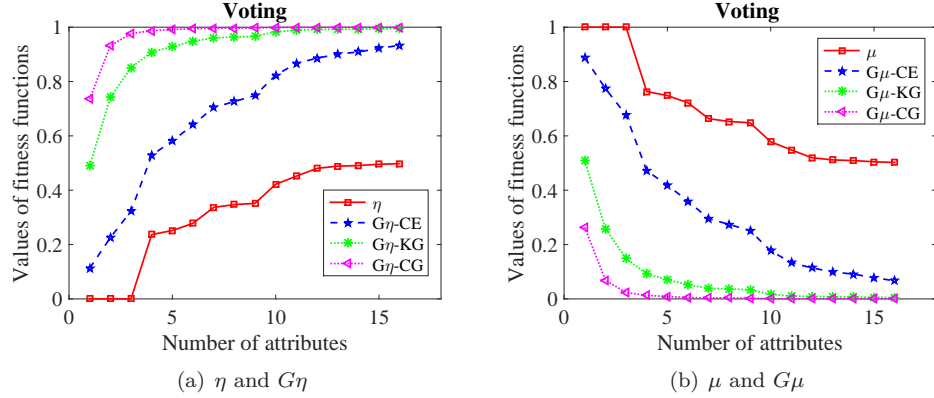


Figure 4: Monotonicity of fitness functions on data set Voting.

Two data sets Vehicle and Credit Approval were used in experiments. There are 18 attributes in Vehicle and 15 attributes in Credit Approval. The experimental results of the fitness functions η , $G\eta - CE$, $G\eta - KG$ and $G\eta - CG$ on two data sets are shown Figures 11 and 13 respectively. The experimental results of the fitness functions μ , $G\mu - CE$, $G\mu - KG$ and $G\mu - CG$ on two data sets are shown Figures 12 and 14 respectively. Similar to the monotonicity experiments, the values of the fitness functions $G\eta - CE$ and $G\mu - CE$ were rescaled to the $[0, 1]$ range.

As to data set Vehicle, for each single attribute, the values of the fitness function η are equal to zero. It indicates that the fitness function η can not evaluate the single attributes effectively. In this situation, we can not rank attributes by the fitness function η . In comparison, the fitness functions $G\eta - CE$, $G\eta - KG$ and $G\eta - CG$ can differentiate the different single attributes

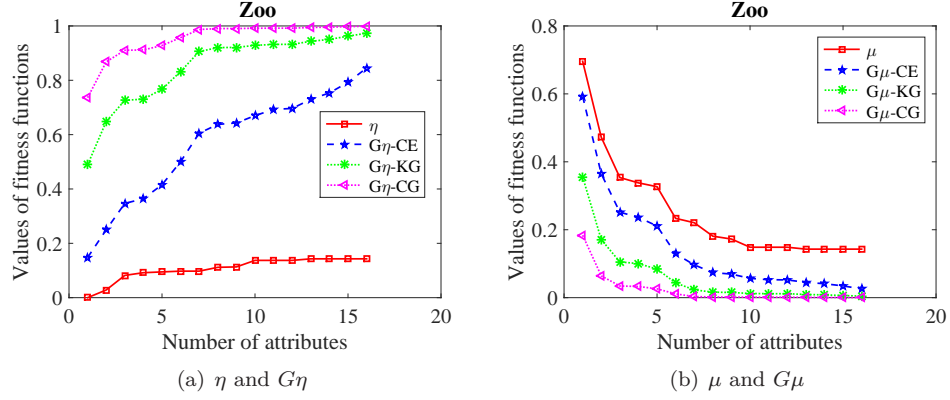


Figure 5: Monotonicity of fitness functions on data set Zoo.

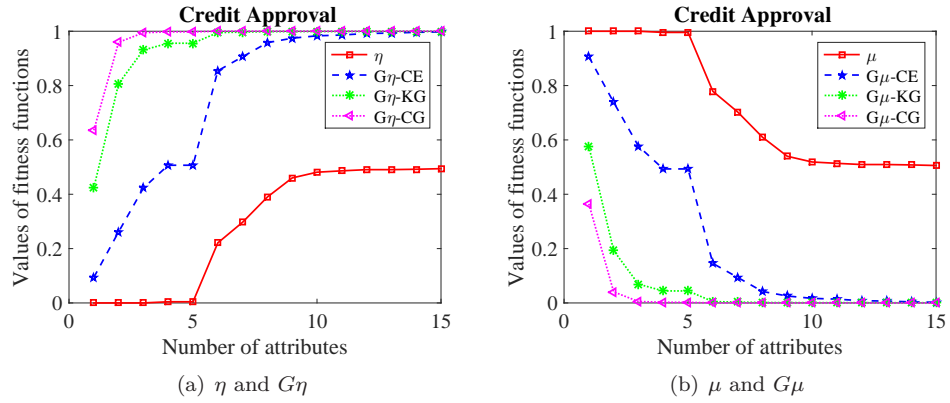


Figure 6: Monotonicity of fitness functions on data set Credit Approval.

effectively. In other words, we can rank attributes effectively by the fitness function $G\eta$. Similarly, for the most single attributes, the values of the fitness function μ are equal to 1. Hence, we can not rank attributes by the fitness function μ , but we can rank attributes by the fitness functions $G\mu - CE$, $G\mu - KG$ and $G\mu - CG$. As to data set Credit Approval, we can obtain a similar result. The above experimental results show that the fitness functions η and μ are not appropriate to evaluate the single attributes and rank them sometimes, and the fitness functions $G\eta$ and $G\mu$ can evaluate the single attributes and rank them effectively.

4.3. The classification accuracy experiments

In this subsection, we took the (α, β) low distribute reduct as an example of the distribute reducts. Then we compared the classification accuracy of

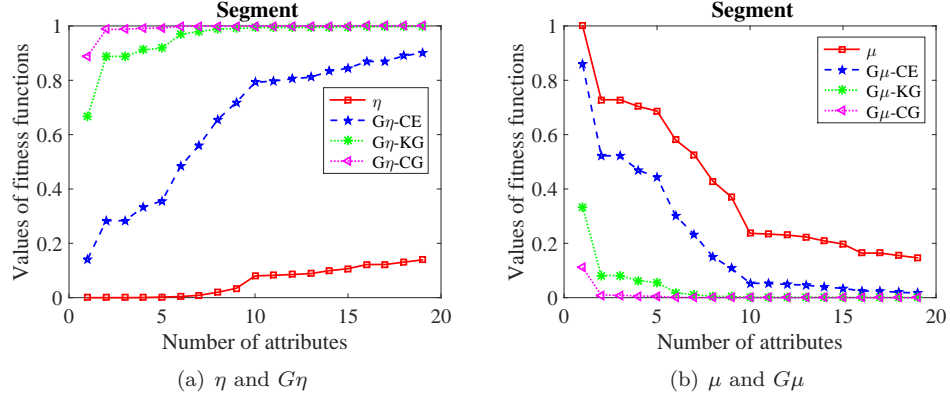


Figure 7: Monotonicity of fitness functions on data set Segment.

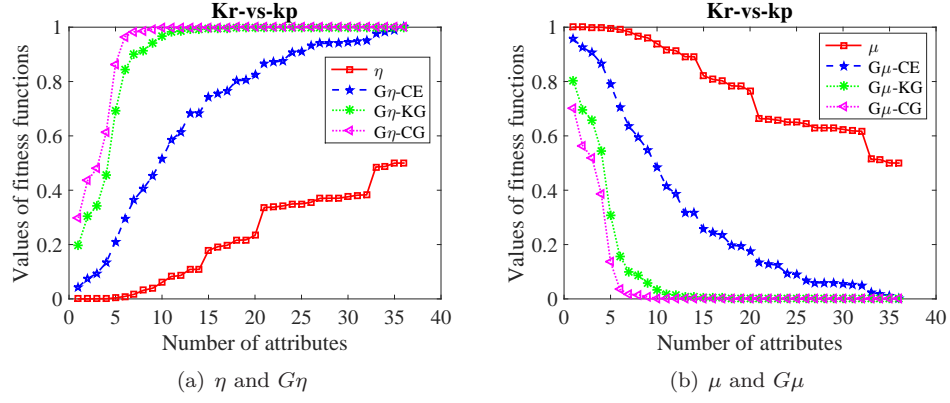


Figure 8: Monotonicity of fitness functions on data set Kr-vs-kp.

the (α, β) low distribution reduct with three traditional definitions of attribute reduct, including the qualitative positive region-preserved reduct (QLPRP) [61], the quantitative positive region-preserved reduct (QNPRP) [61] and the positive region extension reduct (EXDPR) [21], in probabilistic rough set model. For each definition, the corresponding reducts are obtained by the addition-deletion method and the deletion method.

For the (α, β) low distribution reduct, in Algorithms 4 and 5, we considered three different implementations of $G\eta$ by taking CE, KG and CG as the examples of the expected granularity. In addition, for Algorithm 4 and 5, the (α, β) low distribution reducts obtained by three different implementations are denoted as LDRCE, LDRKG and LDRCG respectively.

In the experiments, two well-known used classifiers including BayesNet and linear SVM (SMO) were selected to evaluate the different definitions of attribute

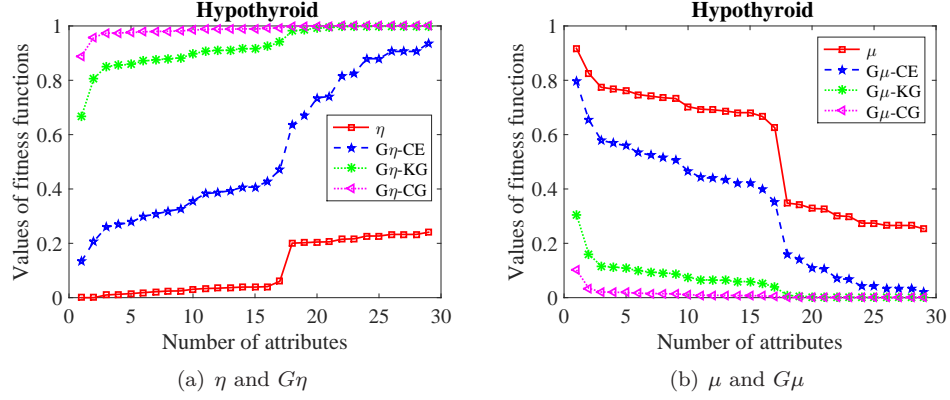


Figure 9: Monotonicity of fitness functions on data set Hypothyroid.

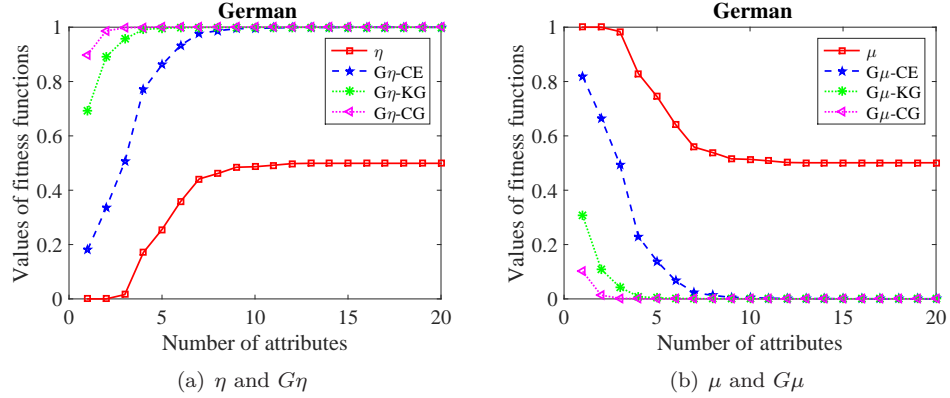


Figure 10: Monotonicity of fitness functions on data set German.

reduct. 10-fold cross-validation scheme was used to assess the performance of the classifiers. All approaches were implemented based on the WEKA data mining software package [9], where the classifiers were implemented with default settings.

To make a fair comparison, we took the value of α from 0.1 to 1.0 with step 0.1. For each data set, the corresponding attribute reducts can be got according to the different α . Then we computed the classification accuracies of all attribute reducts using BayesNet and SMO based on 10-fold cross-validation. The average value and standard deviation were recorded as the final classification accuracies.

Tables 4 - 7 show the experimental results of classification accuracies. In each table, the average maximum classification accuracies are depicted in bold. The classification accuracies are performed on the raw data sets also.

It is observed from Tables 4 - 7 that, for addition-deletion method, LDRCE,

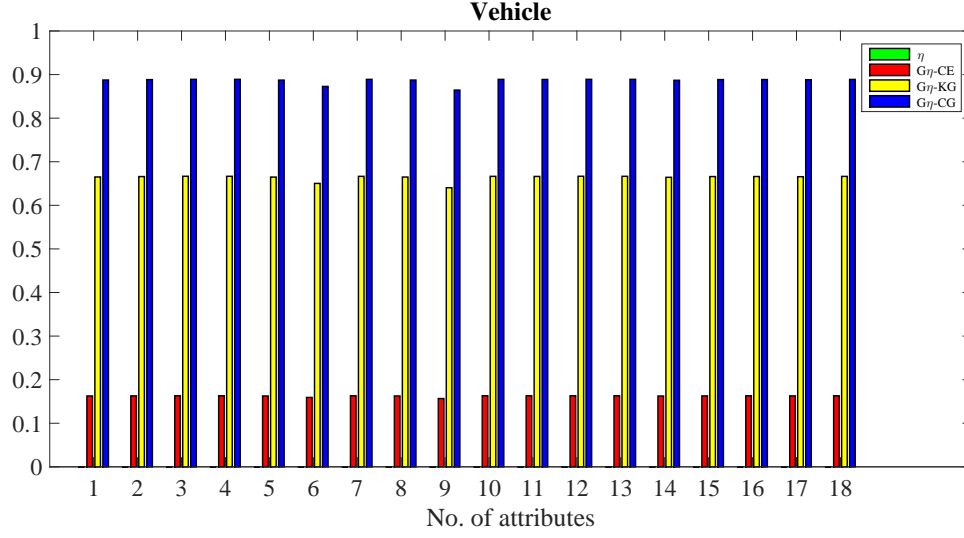


Figure 11: Significance of a single attribute computed with different fitness functions

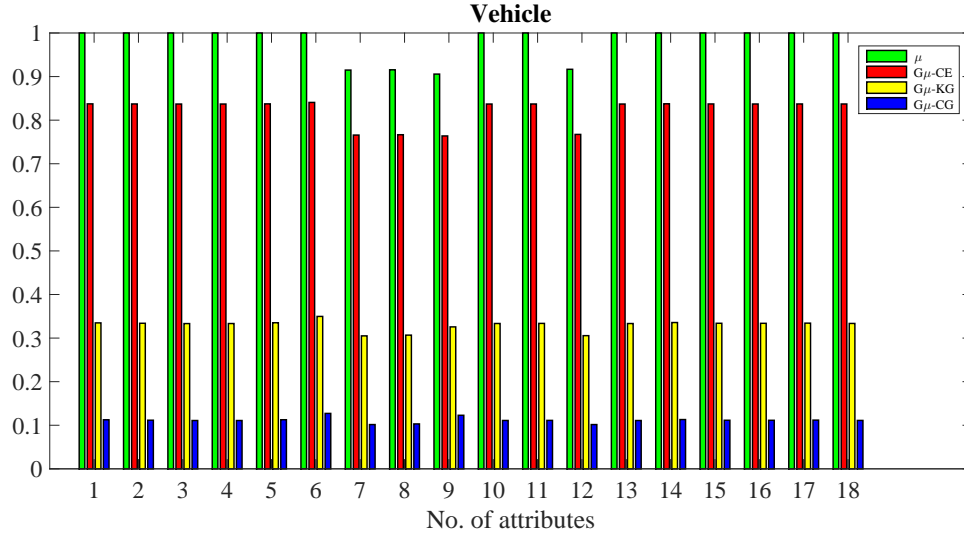


Figure 12: Significance of a single attribute computed with different fitness functions

LDRKG and LDRCG exhibited the best average classification accuracy based on BayesNet in most cases. In detail, LDRCE and LDRCG achieved the highest average classification accuracy on six data sets, and LDRKG achieved the highest average classification accuracy on seven data sets. It should be emphasized that LDRCE, LDRKG and LDRCG jointly achieved the highest average

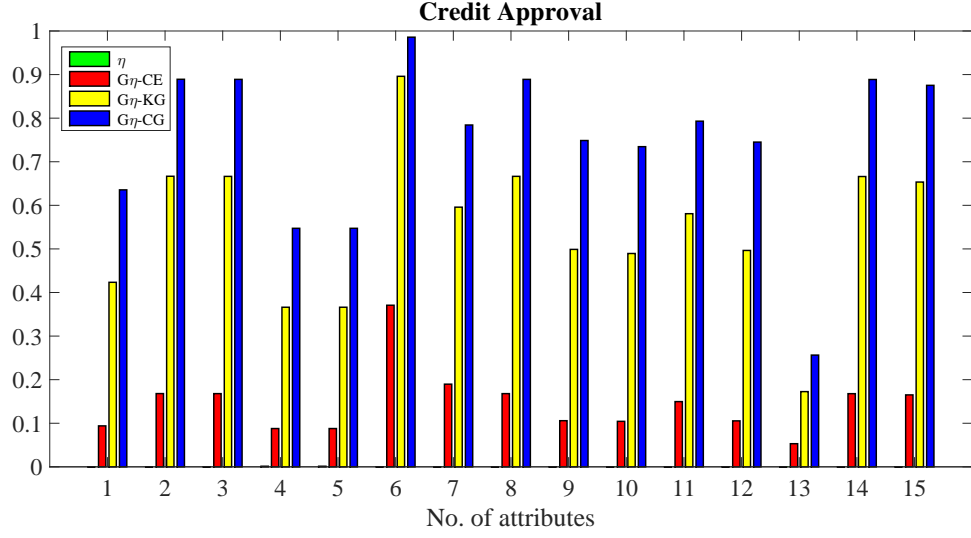


Figure 13: Significance of a single attribute computed with different fitness functions

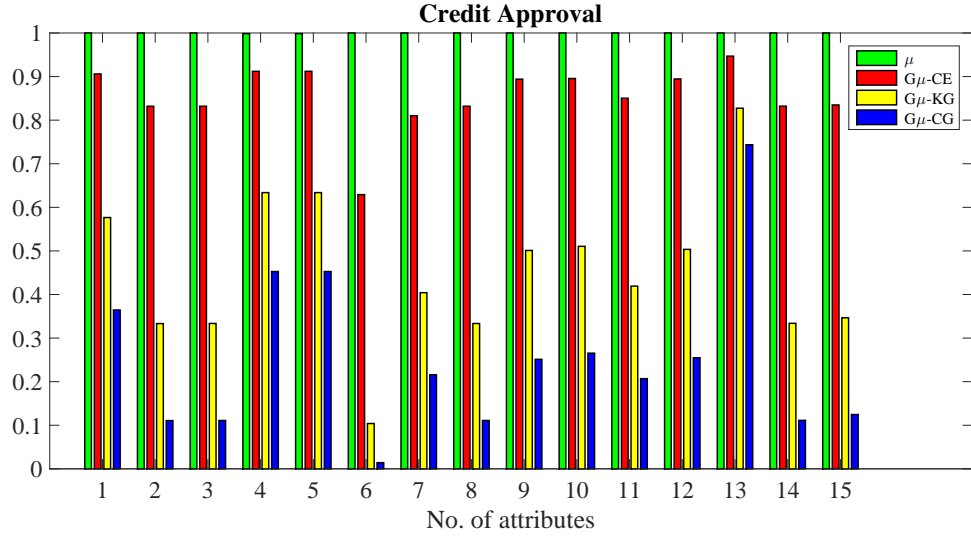


Figure 14: Significance of a single attribute computed with different fitness functions

classification accuracy on five data sets. Moreover, EXDPR got the maximum average classification accuracy on data set Hypothyroid. Compared with the classification accuracy of raw data, LDRCE, LDRKG and LDRCG obtained better classification performance on all the data sets except the data sets Credit Approval and German. In most cases, QLPRPR, QNPRPR and PRER decrease

Table 4: Classification accuracies based on BayesNet by addition-deletion method

ID	RawData	QLPRP	QNPRP	PRER	LDRCE	LDRKG	LDRCG
1	0.8071	0.7174 \pm 0.0910	0.7630 \pm 0.0653	0.8226 \pm 0.0112	0.8234 \pm 0.0000	0.8234 \pm 0.0000	0.8424 \pm 0.0000
2	0.4690	0.4183 \pm 0.0841	0.4174 \pm 0.0860	0.4212 \pm 0.0884	0.4720 \pm 0.0000	0.4720 \pm 0.0000	0.4720 \pm 0.0000
3	0.5201	0.5002 \pm 0.0793	0.4939 \pm 0.0736	0.4852 \pm 0.0751	0.5402 \pm 0.0000	0.5402 \pm 0.0000	0.5402 \pm 0.0000
4	0.9011	0.7657 \pm 0.1520	0.7667 \pm 0.1537	0.7605 \pm 0.1219	0.9195 \pm 0.0000	0.9195 \pm 0.0000	0.9195 \pm 0.0000
5	0.9406	0.7624 \pm 0.1343	0.7881 \pm 0.1549	0.7733 \pm 0.1816	0.9406 \pm 0.0000	0.9505 \pm 0.0000	0.9505 \pm 0.0000
6	0.8623	0.7116 \pm 0.0740	0.7133 \pm 0.0757	0.7086 \pm 0.1535	0.7899 \pm 0.0000	0.7899 \pm 0.0000	0.7899 \pm 0.0000
7	0.8130	0.6715 \pm 0.2357	0.6871 \pm 0.2078	0.6555 \pm 0.2235	0.8433 \pm 0.0000	0.8364 \pm 0.0000	0.8355 \pm 0.0000
8	0.8792	0.7605 \pm 0.1045	0.7239 \pm 0.1694	0.6856 \pm 0.1737	0.8820 \pm 0.0000	0.8820 \pm 0.0000	0.8820 \pm 0.0000
9	0.9176	0.9201 \pm 0.0020	0.9201 \pm 0.0020	0.9224 \pm 0.0014	0.9181 \pm 0.0000	0.9181 \pm 0.0000	0.9181 \pm 0.0000
10	0.7450	0.7165 \pm 0.0165	0.7275 \pm 0.0275	0.7059 \pm 0.0331	0.7360 \pm 0.0000	0.7550 \pm 0.0000	0.7330 \pm 0.0000

Table 5: Classification accuracies based on SMO by addition-deletion method

ID	RawData	QLPRP	QNPRP	PRER	LDRCE	LDRKG	LDRCG
1	0.8071	0.7296 \pm 0.0992	0.7603 \pm 0.0559	0.8212 \pm 0.0102	0.8288 \pm 0.0000	0.8315 \pm 0.0000	0.8288 \pm 0.0000
2	0.4690	0.4121 \pm 0.0855	0.4115 \pm 0.0866	0.4156 \pm 0.0863	0.4720 \pm 0.0000	0.4720 \pm 0.0000	0.4720 \pm 0.0000
3	0.6749	0.5913 \pm 0.1377	0.5914 \pm 0.1377	0.5735 \pm 0.1462	0.6903 \pm 0.0000	0.6903 \pm 0.0000	0.6903 \pm 0.0000
4	0.9586	0.7851 \pm 0.1713	0.7798 \pm 0.1664	0.7641 \pm 0.1271	0.9563 \pm 0.0000	0.9563 \pm 0.0000	0.9563 \pm 0.0000
5	0.9604	0.7713 \pm 0.1447	0.7772 \pm 0.1442	0.7723 \pm 0.1807	0.9406 \pm 0.0000	0.9307 \pm 0.0000	0.9307 \pm 0.0000
6	0.8536	0.7172 \pm 0.0803	0.7248 \pm 0.0871	0.7033 \pm 0.1483	0.8116 \pm 0.0000	0.8116 \pm 0.0000	0.8116 \pm 0.0000
7	0.9203	0.7162 \pm 0.2690	0.7482 \pm 0.2420	0.7106 \pm 0.2610	0.9242 \pm 0.0000	0.9216 \pm 0.0000	0.9212 \pm 0.0000
8	0.9543	0.7896 \pm 0.1389	0.7566 \pm 0.2021	0.7067 \pm 0.1991	0.9549 \pm 0.0000	0.9549 \pm 0.0000	0.9549 \pm 0.0000
9	0.9234	0.9236 \pm 0.0006	0.9236 \pm 0.0006	0.9230 \pm 0.0004	0.9244 \pm 0.0000	0.9244 \pm 0.0000	0.9244 \pm 0.0000
10	0.7440	0.7140 \pm 0.0140	0.7255 \pm 0.0255	0.7078 \pm 0.0348	0.7440 \pm 0.0000	0.7510 \pm 0.0000	0.7280 \pm 0.0000

Table 6: Classification accuracies based on BayesNet by deletion method

ID	RawData	QLPRP	QNPRP	PRER	LDRCE	LDRKG	LDRCG
1	0.7853	0.7242 \pm 0.0950	0.7522 \pm 0.0471	0.8098 \pm 0.0111	0.8288 \pm 0.0000	0.8288 \pm 0.0000	0.8098 \pm 0.0000
2	0.4690	0.4212 \pm 0.0812	0.4227 \pm 0.0797	0.4283 \pm 0.0780	0.4720 \pm 0.0000	0.4720 \pm 0.0000	0.4720 \pm 0.0000
3	0.5201	0.4986 \pm 0.0790	0.4948 \pm 0.0750	0.5030 \pm 0.0735	0.5402 \pm 0.0000	0.5402 \pm 0.0000	0.5402 \pm 0.0000
4	0.9011	0.7720 \pm 0.1582	0.7720 \pm 0.1582	0.7963 \pm 0.1231	0.9310 \pm 0.0000	0.9310 \pm 0.0000	0.9310 \pm 0.0000
5	0.9406	0.7762 \pm 0.1531	0.7762 \pm 0.1531	0.7941 \pm 0.1727	0.9505 \pm 0.0000	0.9505 \pm 0.0000	0.9505 \pm 0.0000
6	0.8623	0.7116 \pm 0.0740	0.7225 \pm 0.0883	0.7123 \pm 0.1573	0.7884 \pm 0.0000	0.7899 \pm 0.0000	0.7899 \pm 0.0000
7	0.8130	0.6885 \pm 0.2304	0.6824 \pm 0.2303	0.6648 \pm 0.2280	0.8355 \pm 0.0000	0.8355 \pm 0.0000	0.8398 \pm 0.0000
8	0.8792	0.7713 \pm 0.1108	0.7239 \pm 0.1694	0.7065 \pm 0.1844	0.8820 \pm 0.0000	0.8820 \pm 0.0000	0.8820 \pm 0.0000
9	0.9176	0.9201 \pm 0.0020	0.9196 \pm 0.0025	0.9194 \pm 0.0039	0.9181 \pm 0.0000	0.9181 \pm 0.0000	0.9181 \pm 0.0000
10	0.7450	0.7180 \pm 0.0180	0.7275 \pm 0.0275	0.7043 \pm 0.0337	0.7360 \pm 0.0000	0.7550 \pm 0.0000	0.7550 \pm 0.0000

the classification accuracies of raw data to some extent. In addition, the classification accuracies of LDRCE, LDRKG and LDRCG was found to be similar

Table 7: Classification accuracies based on SMO by deletion method

ID	RawData	QLPRP	QNPRP	PRER	LDRCE	LDRKG	LDRCG
1	0.8071	0.7174 \pm 0.0910	0.7630 \pm 0.0653	0.8226 \pm 0.0112	0.8234 \pm 0.0000	0.8234 \pm 0.0000	0.8424 \pm 0.0000
2	0.4690	0.4156 \pm 0.0813	0.4174 \pm 0.0791	0.4224 \pm 0.0770	0.4720 \pm 0.0000	0.4720 \pm 0.0000	0.4720 \pm 0.0000
3	0.6749	0.5894 \pm 0.1376	0.5882 \pm 0.1370	0.6046 \pm 0.1270	0.6903 \pm 0.0000	0.6903 \pm 0.0000	0.6903 \pm 0.0000
4	0.9586	0.7841 \pm 0.1703	0.7841 \pm 0.1703	0.8145 \pm 0.1409	0.9540 \pm 0.0000	0.9540 \pm 0.0000	0.9540 \pm 0.0000
5	0.9604	0.7752 \pm 0.1518	0.7752 \pm 0.1518	0.7931 \pm 0.1719	0.9406 \pm 0.0000	0.9406 \pm 0.0000	0.9406 \pm 0.0000
6	0.8536	0.7172 \pm 0.0803	0.7288 \pm 0.0919	0.7036 \pm 0.1486	0.8043 \pm 0.0000	0.8116 \pm 0.0000	0.8116 \pm 0.0000
7	0.9203	0.7485 \pm 0.2548	0.7309 \pm 0.2633	0.7246 \pm 0.2651	0.9212 \pm 0.0000	0.9212 \pm 0.0000	0.9242 \pm 0.0000
8	0.9543	0.8077 \pm 0.1472	0.7566 \pm 0.2021	0.7395 \pm 0.2173	0.9549 \pm 0.0000	0.9549 \pm 0.0000	0.9549 \pm 0.0000
9	0.9234	0.9234 \pm 0.0005	0.9233 \pm 0.0008	0.9233 \pm 0.0005	0.9242 \pm 0.0000	0.9242 \pm 0.0000	0.9242 \pm 0.0000
10	0.7440	0.7220 \pm 0.0220	0.7255 \pm 0.0255	0.7078 \pm 0.0348	0.7440 \pm 0.0000	0.7510 \pm 0.0000	0.7510 \pm 0.0000

or consistent. As to SMO, LDRCE, LDRKG and LDRCG have better average classification accuracy than QLPRP, QNPRP and PRER on all the data sets. Compared with the classification accuracy of raw data, LDRCE, LDRKG and LDRCG obtained better classification performance on more than half of data sets. For the deletion method, the similar results can be obtained by the classification accuracy based on BayesNet and SMO.

Furthermore, from Tables 4 - 7, we can clearly notice that the standard deviations of classification accuracies derived from LDRCE, LDRKG and LDRCG equal to zero on all the data sets. It implies Algorithms 4 and 5 are insensitive to threshold parameters. This can be explained because the (α, β) low distribution reduct has a more stringent reduction condition that preserves the (α, β) low approximation of all decision classes. In fact, the (α, β) low distribution reduct is equivalent to reduct that keeps Pawlak positive region (Definition 2) unchanged when decision table is consistent ($POS_R(D) = U$). It means that for the consistent decision table, the classification accuracies derived from LDRCE, LDRKG and LDRCG are the same for all threshold parameters. What's more, for inconsistent decision table ($POS_R(D) \neq U$), Algorithms 4 and 5 are also insensitive to threshold parameters when each decision class contains too few inconsistent objects because the (α, β) low distribution reduct must keep the (α, β) low approximation of all decision classes unchanged.

Table 8 and 9 outlined the average length of the derived reduct based on the addition-deletion method and the deletion method. From Table 8 and 9, it can be seen that the numbers of attributes obtained by QLPRP, QNPRP and PRER are less than those obtained by LDRCE, LDRKG and LDRCG. The reason should be attributed to the fact that the (α, β) low distribution reduct have more stringent reduction condition than three other definitions of attribute reduct. For length of the derived reduct, the standard deviations of LDRCE, LDRKG and LDRCG also equal to zero on all the data sets. It further indicates Algorithms 4 and 5 are insensitive to threshold parameters. In addition, for QLPRP, QNPRP and PRER, the relatively high standard deviations of reduct length also show that QLPRP, QNPRP and PRER

obtained by the addition-deletion method and the deletion method are sensitive to threshold parameters.

Table 8: Average length of a reduct based on the addition-deletion method

ID	QLPRP	QNPRP	PRER	LDRCE	LDRKG	LDRCG
1	5.5 \pm 4.5	4.6 \pm 4.4	3.7 \pm 4.1	10.0 \pm 0.0	10.0 \pm 0.0	10.0 \pm 0.0
2	11.9 \pm 5.7	11.9 \pm 5.7	11.9 \pm 5.6	16.0 \pm 0.0	16.0 \pm 0.0	16.0 \pm 0.0
3	9.5 \pm 6.1	9.4 \pm 6.1	8.6 \pm 6.5	15.0 \pm 0.0	15.0 \pm 0.0	15.0 \pm 0.0
4	6.2 \pm 5.3	6.6 \pm 5.6	2.1 \pm 3.3	12.0 \pm 0.0	12.0 \pm 0.0	12.0 \pm 0.0
5	3.8 \pm 2.9	3.4 \pm 2.4	3.0 \pm 2.0	7.0 \pm 0.0	6.0 \pm 0.0	6.0 \pm 0.0
6	6.0 \pm 5.0	5.9 \pm 4.9	4.0 \pm 4.6	11.0 \pm 0.0	11.0 \pm 0.0	11.0 \pm 0.0
7	8.6 \pm 6.4	9.4 \pm 5.1	7.6 \pm 5.3	15.0 \pm 0.0	14.0 \pm 0.0	14.0 \pm 0.0
8	12.3 \pm 13.6	15.0 \pm 14.0	9.6 \pm 12.7	29.0 \pm 0.0	29.0 \pm 0.0	29.0 \pm 0.0
9	14.5 \pm 9.5	14.2 \pm 9.3	3.2 \pm 6.6	23.0 \pm 0.0	23.0 \pm 0.0	23.0 \pm 0.0
10	5.5 \pm 4.5	5.5 \pm 4.5	4.9 \pm 4.8	10.0 \pm 0.0	10.0 \pm 0.0	10.0 \pm 0.0

Table 9: Average length of a reduct based on the deletion method

ID	QLPRP	QNPRP	PRER	LDRCE	LDRKG	LDRCG
1	5.1 \pm 4.1	5.5 \pm 4.5	5.5 \pm 4.5	9.0 \pm 0.0	9.0 \pm 0.0	10.0 \pm 0.0
2	12.0 \pm 5.6	12.1 \pm 5.5	12.5 \pm 5.2	16.0 \pm 0.0	16.0 \pm 0.0	16.0 \pm 0.0
3	9.6 \pm 6.1	9.4 \pm 6.1	9.8 \pm 5.6	15.0 \pm 0.0	15.0 \pm 0.0	15.0 \pm 0.0
4	6.2 \pm 5.3	6.2 \pm 5.3	6.3 \pm 5.3	12.0 \pm 0.0	12.0 \pm 0.0	12.0 \pm 0.0
5	2.9 \pm 2.1	2.9 \pm 2.1	3.2 \pm 1.9	6.0 \pm 0.0	6.0 \pm 0.0	6.0 \pm 0.0
6	6.0 \pm 5.0	6.0 \pm 5.0	5.9 \pm 4.9	11.0 \pm 0.0	11.0 \pm 0.0	11.0 \pm 0.0
7	8.7 \pm 5.6	8.7 \pm 6.2	8.2 \pm 5.2	14.0 \pm 0.0	14.0 \pm 0.0	14.0 \pm 0.0
8	15.0 \pm 14.0	15.0 \pm 14.0	15.0 \pm 14.0	29.0 \pm 0.0	29.0 \pm 0.0	29.0 \pm 0.0
9	14.5 \pm 9.5	14.1 \pm 9.3	12.6 \pm 9.7	23.0 \pm 0.0	23.0 \pm 0.0	23.0 \pm 0.0
10	5.5 \pm 4.5	5.5 \pm 4.5	6.0 \pm 5.1	10.0 \pm 0.0	10.0 \pm 0.0	10.0 \pm 0.0

From the experimental results, we can draw a conclusion that the distribution reducts are relatively the better choices in probabilistic rough set model, as attributes they selected have the higher classification accuracy.

4.4. Comparison of distribution reducts with ranking based attribute reduction methods

In this subsection, we compared the classification accuracy of the distribution reducts with ranking based attribute reduction methods (RBAR). RBAR is a classic attribute reduction algorithm. In RBAR, the attributes are globally ranked based on the significance of single attributes, and the first k best attributes are selected. In our experiments, the significance of single attributes was evaluated by using the fitness functions $G\eta - CE$, $G\eta - KG$ and $G\eta - CG$, and the different values of k were specified based on the subset size of the corresponding (α, β) low distribution reducts, which are obtained by the size of LDRCE, LDRKG and LDRCG, to compare the distribution reducts with

Table 10: Classification accuracies based on BayesNet by ranking method (subset size obtained by addition-deletion method)

ID	RBARCE	RBARKG	RBARCG
1	0.7745 ± 0.0000	0.7147 ± 0.0000	0.7147 ± 0.0000
2	0.4720 ± 0.0000	0.4720 ± 0.0000	0.4720 ± 0.0000
3	0.4504 ± 0.0000	0.4504 ± 0.0000	0.4504 ± 0.0000
4	0.8943 ± 0.0000	0.8943 ± 0.0000	0.8943 ± 0.0000
5	0.9307 ± 0.0000	0.8416 ± 0.0000	0.8416 ± 0.0000
6	0.8652 ± 0.0000	0.8652 ± 0.0000	0.8652 ± 0.0000
7	0.8260 ± 0.0000	0.8312 ± 0.0000	0.8212 ± 0.0000
8	0.8767 ± 0.0000	0.8767 ± 0.0000	0.8767 ± 0.0000
9	0.9178 ± 0.0000	0.9178 ± 0.0000	0.9223 ± 0.0000
10	0.7430 ± 0.0000	0.7420 ± 0.0000	0.7420 ± 0.0000

ranking based attribute reduction methods. The classification accuracies were obtained by adopting the same experimental settings and methods that were used in subsection 4.3.

Tables 10-13 present experimental results of RBAR on ten data sets using BayesNet and SMO, where RBARCE, RBARKG and RBARCG represent the attribute reducts obtained by using $G\eta-CE$, $G\eta-KG$ and $G\eta-CG$ to evaluate the significance of single attributes, respectively. Compare with Tables 4 - 7, it is easy to see that, for the (α, β) low distribution reducts obtained by the addition-deletion method, LDRCE and LDRCG outperform RBARCE and RBARCG on seven data sets respectively, and LDRKG outperform RBARKG on eight data sets according to BayesNet. Note that LDRCE, LDRKG and LDRCG get the same average classification accuracy on data set Primary-tumor with RBARCE, RBARKG and RBARCG respectively. This is because the (α, β) low distribution reducts obtained by the addition-deletion method and reducts obtained by ranking method are the same. As to SMO, the results is similar to that of BayesNet. For the (α, β) low distribution reducts obtained by the deletion method, we can obtain the similar results also.

Nevertheless, from Table 8 and 9, we can easily find that the standard deviations of classification accuracies derived from RBARCE, RBARKG and RBARCG equal to zero on all the data sets. This is because, for given 10 different values of α , we always obtain the same ranking of attributes according to the fitness function $G\eta$, and the subset size of the (α, β) low distribution reduct obtained by the addition-deletion method or deletion method is same every time. It further indicates that Algorithms 4 and 5 are insensitive to threshold parameters.

The above experimental results show that the distribution reducts are effective.

Table 11: Classification accuracies based on SMO by ranking method (subset size obtained by addition-deletion method)

ID	RBARCE	RBARKG	RBARCG
1	0.7391 ± 0.0000	0.7147 ± 0.0000	0.7147 ± 0.0000
2	0.4720 ± 0.0000	0.4720 ± 0.0000	0.4720 ± 0.0000
3	0.6040 ± 0.0000	0.6040 ± 0.0000	0.6040 ± 0.0000
4	0.9563 ± 0.0000	0.9563 ± 0.0000	0.9563 ± 0.0000
5	0.9307 ± 0.0000	0.8416 ± 0.0000	0.8416 ± 0.0000
6	0.8551 ± 0.0000	0.8551 ± 0.0000	0.8551 ± 0.0000
7	0.9216 ± 0.0000	0.9203 ± 0.0000	0.9048 ± 0.0000
8	0.9528 ± 0.0000	0.9528 ± 0.0000	0.9528 ± 0.0000
9	0.9236 ± 0.0000	0.9236 ± 0.0000	0.9229 ± 0.0000
10	0.7490 ± 0.0000	0.7520 ± 0.0000	0.7520 ± 0.0000

Table 12: Classification accuracies based on BayesNet by ranking method(subset size obtained by deletion method)

ID	RBARCE	RBARKG	RBARCG
1	0.7663 ± 0.0000	0.7092 ± 0.0000	0.7147 ± 0.0000
2	0.4720 ± 0.0000	0.4720 ± 0.0000	0.4720 ± 0.0000
3	0.4504 ± 0.0000	0.4504 ± 0.0000	0.4504 ± 0.0000
4	0.8943 ± 0.0000	0.8943 ± 0.0000	0.8943 ± 0.0000
5	0.8416 ± 0.0000	0.8416 ± 0.0000	0.8416 ± 0.0000
6	0.8652 ± 0.0000	0.8652 ± 0.0000	0.8652 ± 0.0000
7	0.8312 ± 0.0000	0.8312 ± 0.0000	0.8087 ± 0.0000
8	0.8767 ± 0.0000	0.8767 ± 0.0000	0.8767 ± 0.0000
9	0.9178 ± 0.0000	0.9178 ± 0.0000	0.9223 ± 0.0000
10	0.7430 ± 0.0000	0.7420 ± 0.0000	0.7420 ± 0.0000

5. Conclusion

The addition-deletion method based and deletion method based attribute reduction algorithms are two representative heuristic attribute reduction algorithms in rough set theory. The fitness functions play a crucial role in designing heuristic attribute reduction algorithms. The monotonicity of fitness functions is very important to guarantee the validity of heuristic attribute reduction algorithms. This paper aims at developing heuristic attribute reduction algorithms for finding distribution reducts in probabilistic rough set model. To begin with, we proposed two monotonic fitness functions $\eta_R^{(\alpha, \beta)}$ and $\mu_R^{(\alpha, \beta)}$. The equivalence definitions of the (α, β) low and upper distribution reducts are given respectively based on $\eta_R^{(\alpha, \beta)}$ and $\mu_R^{(\alpha, \beta)}$. The equivalence definitions of two distribution reducts can be use to design the stopping criteria of heuristic attribute

Table 13: Classification accuracies based on SMO by ranking method(subset size obtained by deletion method)

ID	RBARCE	RBARKG	RBARCG
1	0.7527 ± 0.0000	0.6902 ± 0.0000	0.7147 ± 0.0000
2	0.4720 ± 0.0000	0.4720 ± 0.0000	0.4720 ± 0.0000
3	0.6040 ± 0.0000	0.6040 ± 0.0000	0.6040 ± 0.0000
4	0.9563 ± 0.0000	0.9563 ± 0.0000	0.9563 ± 0.0000
5	0.8416 ± 0.0000	0.8416 ± 0.0000	0.8416 ± 0.0000
6	0.8551 ± 0.0000	0.8551 ± 0.0000	0.8551 ± 0.0000
7	0.9203 ± 0.0000	0.9203 ± 0.0000	0.9078 ± 0.0000
8	0.9528 ± 0.0000	0.9528 ± 0.0000	0.9528 ± 0.0000
9	0.9236 ± 0.0000	0.9236 ± 0.0000	0.9229 ± 0.0000
10	0.7490 ± 0.0000	0.7520 ± 0.0000	0.7520 ± 0.0000

reduction algorithms. Additionally, we presented two modified fitness functions $G\eta_R^{(\alpha,\beta)}$ and $G\mu_R^{(\alpha,\beta)}$ by multiplying measures of granularity of partitions by $\eta_R^{(\alpha,\beta)}$ and $\mu_R^{(\alpha,\beta)}$ to evaluate the significance of attributes more effectively. On this basis, we developed two heuristic attribute reduction algorithms to find distribution reducts based on the addition-deletion method and deletion method. The monotonicity of fitness functions proposed guarantees that the reduction result is right as well. Finally, experiments on several real-life data sets are conducted to test the effectiveness of the proposed fitness functions and distribution reducts. The experimental results show that $G\eta_R^{(\alpha,\beta)}$ and $G\mu_R^{(\alpha,\beta)}$ are a more effective alternative than $\eta_R^{(\alpha,\beta)}$ and $\mu_R^{(\alpha,\beta)}$ for evaluating the significance of attributes. The experimental results also indicate that distribution reducts can achieve better performance in probabilistic rough set model and outperforms ranking based attribute reduction methods.

This investigation provides a new insight into the distribution reducts in probabilistic rough set model. In the future, the study work will be focused on designing heuristic attribute reduction algorithms to find other definitions of attribute reduct in probabilistic rough set model by studying monotonic fitness functions.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant Nos. 61502419, 61272060, 61379114), and the Key Natural Science Foundation of Chongqing (No. CSTC2013jjB40003).

References

- [1] Ucirvine machine learning repository, <http://archive.ics.uci.edu/ml/>.

- [2] D. Bianucci and G. Cattaneo. Information entropy and granulation co-entropy of partitions and coverings: A summary. *Transactions on Rough Sets X*, 5656:15–66, 2009.
- [3] G. Cattaneo, D. Ciucci, and D. Bianucci. Entropy and co-entropy of partitions and coverings with applications to roughness theory. *Granular Computing: At the Junction of Rough Sets and Fuzzy Sets*, 224:55–77, 2008.
- [4] Y. M. Chen, D. Q. Miao, and R. Z. Wang. A rough set approach to feature selection based on ant colony optimization. *Pattern Recognition Letters*, 31:226–233, 2010.
- [5] J. H. Dai, W. T. Wang, and Q. Xu. An uncertainty measure for incomplete decision tables and its applications. *IEEE Transactions on Cybernetics*, 43(4):1277–1289, 2012.
- [6] J. H. Dai and Q. Xu. Approximations and uncertainty measures in incomplete information systems. *Information Sciences*, 198:62–80, 2012.
- [7] T. Q. Deng, C. D. Yang, and Q. H. Hu. Feature selection in decision systems based on conditional knowledge granularity. *International Journal of Computational intelligence systems*, 4(4):655–671, 2011.
- [8] H. L. Dou, X. B. Yang, X. N. Song, H. L. Yu, W. Z. Wu, and J. Y. Yang. Decision-theoretic rough set: A multicost strategy. *Knowledge-Based Systems*, 000:1–13, 2015.
- [9] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: an update, sigkdd explorations. 11(1):10–18, 2009.
- [10] T. Herawan, M. M. Deris, and J. H. Abawajy. A rough set approach for selecting clustering attribute. *Knowledge-Based Systems*, 23(3):220–231, 2010.
- [11] J. P. Herbert and J. T. Yao. Criteria for choosing a rough set model. *Computers and Mathematics with applications*, 57(6):908–918, 2009.
- [12] J. P. Herbert and J. T. Yao. Game-theoretic rough sets. *Fundamenta Informaticae*, 108(3):267–286, 2011.
- [13] M. Inuiguchi. Several approaches to attribute reduction in variable precision rough set model. *Modeling Decisions for Artificial Intelligence*, pages 215–226, 2005.
- [14] X. Y. Jia, W. H. Liao, Z. M. Tang, and L. Shang. Minimum cost attribute reduction in decision-theoretic rough set models. *Information Sciences*, 219(4):151C167, 2013.
- [15] X. Y. Jia, L. Shang, B. Zhou, and Y. Y. Yao. Generalized attribute reduct in rough set theory. *Knowledge-Based Systems*, 51(4):453–471, 2015.

- [16] X. Y. Jia, Tang, W. H. Z. M. Liao, and L. Shang. On an optimization representation of decision-theoretic rough set model. *International Journal Of Approximate Reasoning*, 55(1):156–166, 2014.
- [17] F. Jiang, Y. F. Sui, and L. Zhou. A relative decision entropy-based feature selection approach. *Pattern Recognition*, 48:2151–2163, 2015.
- [18] M. Kryszkiewicz. Comparative study of alternative types of knowledge reduction in inconsistent systems. *International Journal of Intelligence Systems*, 16(1):105–120, 2001.
- [19] M. Kryszkiewicz. Certain, generalized decision, and membership distribution reducts versus functional dependencies in incomplete systems. In *Proceeding of RSEISP2007, LNAI 4585*, pages 162–174, 2007.
- [20] F. C. Li, Z. Zhang, and C. X. Jin. Constructing importance measure of attributes in covering decision table. *Knowledge-Based Systems*, 76:228–239, 2015.
- [21] H. X. Li, X. Z. Zhou, J. B. Zhao, and D. Liu. Non-monotonic attribute reduction in decision-theoretic rough sets. *Fundamenta Informaticae*, 126(4):415–432, 2013.
- [22] J. Y. Liang and Z. Z. Shi. The information entropy, rough entropy and knowledge granulation in rough set theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 12(1):37–46, 2004.
- [23] J. Y. Liang, J. H. Wang, and Y. H. Qian. A new measure of uncertainty based on knowledge granulation for rough sets. *Information Sciences*, 179(4):458–470, 2009.
- [24] T. Y. Lin and Y. R. Syau. Unifying variable precision and classical rough sets: Granular approach. In *Rough Sets and Intelligent Systems-Professor Zdzisław Pawlak in Memoriam*, pages 365–373. Springer, 2013.
- [25] D. Liu, T. R. Li, and H. X. Li. A multiple-category classification approach with decision-theoretic rough sets. *Fundamenta Informaticae*, 115(2):173–188, 2012.
- [26] D. Liu, T. R. Li, and D. Ruan. Probabilistic model criteria with decision-theoretic rough sets. *Information Sciences*, 181(17):173–178, 2011.
- [27] J. N. K. Liu, Y. X. Hu, and Y. L. He. A set covering based approach to find the reduct of variable precision rough set. *Information Sciences*, 275(17):83–100, 2014.
- [28] W. M. Ma and B. Z. Sun. Probabilistic rough set over two universes and rough entropy. *International Journal of Approximate Reasoning*, 53(4):608–619, 2012.

- [29] X. A. Ma, G. Y. Wang, and H. Yu. Decision region distribution preservation reduction in decision-theoretic rough set model. *Information Sciences*, 278:614–640, 2014.
- [30] J. S. Mi, W. Z. Wu, and W. X. Zhang. Approaches to knowledge reduction based on variable precision rough set model. *Information Sciences*, 159(3):255–272, 2004.
- [31] D. Q. Miao and G. R. Hu. A heuristic algorithm for reduction of knowledge. *Chinese Journal of Computer Research & Development*, 36(6):681–684, 1999.
- [32] D. Q. Miao, Y. Zhao, H. X. Li, and F. F. Xu. Relative reducts in consistent and inconsistent decision tables of the Pawlak rough set model. *Information Sciences*, 179(24):4140–4150, 2009.
- [33] F. Min, H. P. He, Y. H. Qian, and W. Zhu. Test-cost-sensitive attribute reduction. *Information Sciences*, 181(22):4928–4942, 2011.
- [34] F. Min and W. Zhu. Attribute reduction of data with error ranges and test costs. *Information Sciences*, 211:48–67, 2012.
- [35] N. M. Parthalaian and Q. Shen. Exploring the boundary region of tolerance rough sets for feature selection. *Pattern Recognition*, 42(5):655–667, 2009.
- [36] N. M. Parthalaian, Q. Shen, and R. Jensen. A distance measure approach to exploring the rough set boundary region for attribute reduction. *IEEE Transactions on Knowledge and Data Engineering*, 22(3):305–317, 2010.
- [37] Z. Pawlak. Rough sets. *International Journal of Computer and Information Science*, 11(5):341–356, 1982.
- [38] Z. Pawlak. *Rough sets: theoretical aspects of reasoning about data*. Kluwer Academic Publishers, Boston, 1991.
- [39] Z. Pawlak, S. K. M. Wong, and W. Ziarko. Rough sets: probabilistic versus deterministic approach. *International Journal of Man-Machine Studies*, 29(1):81–95, 1988.
- [40] W. Pedrycz. *Granular Computing: Analysis and Design of Intelligent Systems*. CRC Press/Francis Taylor, Boca Raton, 2013.
- [41] Y. H. Qian and J. Y. Liang. Combination entropy and combination granulation in rough set theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 16(2):179–193, 2008.
- [42] Y. H. Qian, J. Y. Liang, W. Pedrycz, and C. Y. Dang. Positive approximation: An accelerator for attribute reduction in rough set theory. *Artificial Intelligence*, 174(9):597–618, 2010.

- [43] Chebrolu S. and Sanjeevi S. G. Attribute reduction in decision-theoretic rough set models using genetic algorithm. In *Proceeding of 2th International Conference on the Swarm, Evolutionary, and Memetic Computing. LNCS 7076*, pages 307–314, 2011.
- [44] Q. Shen and R. Jensen. Rough sets, their extensions and applications. *International Journal of Automation and Computing*, 4(3):217–228, 2007.
- [45] W. H. Shu and W. B. Qian. An incremental approach to attribute reduction from dynamic incomplete decision systems in rough set theory. *Data & Knowledge Engineering*, 100:116 – 132, 2015.
- [46] D. Slezak and W. Ziarko. The investigation of the bayesian rough set model. *International Journal Approximate Reasoning*, 40(1):81–89, 2005.
- [47] S. H. Teng, M. Lu, A. F. Yang, J. Zhang, Y. J. Nian, and M. He. Efficient attribute reduction from the viewpoint of discernibility. *Information Sciences*, 326:297–314, 2016.
- [48] K. Thangavela and A. Pethalakshmi. Dimensionality reduction based on rough set theory: A review. *Applied Soft Computing*, 9(1):1–12, 2009.
- [49] C. Y. Wang and B. Q. Hu. Granular variable precision fuzzy rough sets with general fuzzy relations. *Fuzzy Sets and Systems*, 275:39–57, 2015.
- [50] J. Y. Wang and J. Zhou. Research of reduct features in the variable precision rough set model. *Neurocomputing*, 72(10):2643–2648, 2009.
- [51] Y. Y. Yang, D. G. Chen, and Z. Dong. Novel algorithms of attribute reduction with variable precision rough set model. *Neurocomputing*, 139:336–344, 2014.
- [52] Y. Y. Yao, S. K. M. Wong, and P. Lingras. A decision-theoretic rough set model. In M. Emrich M.L. (Ed) Ras, Z.W. Zemankova, editor, *Methodologies for Intelligent Systems 5*.
- [53] Y. Y. Yao and L. Z. Zhao. A measurement theory view on the granularity of partitions. *Information Sciences*, 213:1–13, 2012.
- [54] Y. Y. Yao and Y. Zhao. Attribute reduction in decision-theoretic rough set model. *Information Sciences*, 178(17):3356–3373, 2008.
- [55] Y. Y. Yao and Y. Zhao. Discernibility matrix simplification for constructing attribute reducts. *Information Sciences*, 179(7):867–882, 2009.
- [56] Y. Y. Yao, Y. Zhao, and J. Wang. On reduct construction algorithms. In *Transactions on Computational Science II*, volume 5150, pages 100–117, 2008.

- [57] D. Y. Ye, Z. J. Chen, and S. L. Ma. A novel and better fitness evaluation for rough set based minimum attribute reduction problem. *Information Sciences*, 222:413–423, 2013.
- [58] W. W. Zhang, J. S. Mi, and W. Z. Wu. Approaches to knowledge reductions in inconsistent systems. *International Journal of Intelligent Systems*, 18(9):989–1000, 2003.
- [59] X. Y. Zhang and D. Q. Miao. Region-based quantitative and hierarchical attribute reduction in the two-category decision theoretic rough set model. *Knowledge-Based Systems*, 71:146–161, 2014.
- [60] Y. Zhao, F. Luo, S. K. M. Wong, and Y. Y. Yao. A general definition of an attribute reduct. In *Proceeding of RSKT2007*, volume 4481, pages 101–108, 2007.
- [61] Y. Zhao, S. K. M. Wong, and Y. Y. Yao. A note on attribute reduction in the decision-theoretic rough set model. *Transactions on Rough Sets XIII*, 6499:61–70, 2011.
- [62] K. Zheng, J. Hu, Z. F. Zhan, J. Ma, and Qi. J. Jin. An enhancement for heuristic attribute reduction algorithm in rough set. 41(15):6748–6754, 2014.
- [63] W. Ziarko. Variable precision rough set model. *Journal of Computer and System Sciences*, 46(1):39–59, 1993.